United Nations E/CN.3/1999/8

Economic and Social Council

Distr.: General
16 November 1998

Original: English

Statistical Commission
Thirtieth session
1–5 March 1999
Item 3 (e) of the provisional agenda*
Economic statistics: other economic statistics

Evaluation of the International Comparison Programme

Note by the Secretary-General

The Secretary-General has the honour to transmit to the Statistical Commission the report of the consultant on the evaluation of the International Comparison Programme (ICP), which is contained in the annex. The report was prepared in accordance with the request of the Statistical Commission at its twenty-ninth session,[1] under the sponsorship of the World Bank, the International Monetary Fund and the United Nations.

* E/CN.3/1999/1.

[1] See *Official Records of the Economic and Social Council, 1997, Supplement No. 4* (E/1997/24), para. 13 (j)–(l).

98-36798 (E) 010299

Annex

Report of the consultant on the evaluation of the International Comparison Programme*

Contents

		Paragraphs	Page
	Executive summary	1–21	4
	Why this report?	2–6	4
	Past criticisms and past reactions	7–11	6
	Are PPPs required?	12–17	7
	Are ICP estimates credible?	18–20	8
	Recommendations	21	9
	Conclusions	22	10
I.	Introduction	23–30	11
II.	A review of ICP	31–45	12
	What is the subject of ICP?	31–37	12
	How international comparisons enter into everyday discourse	38–41	14
	Why was this report commissioned?	42–45	15
III.	What this report is and what it is not: existing literature and perceived problems	46–50	17
IV.	Why the "unease" with ICP?	51–65	18
	Is ICP justified in view of its applications?	51–56	18
	Is ICP justified in view of the quality of data it is likely to produce?	57	22
	Is it better to have no PPP data than data produced by ICP?	58–59	23
	What is the minimum investment required for a substantial improvement in the quality of PPP data	60–62	23
	How would we know better-quality data if we saw them?	63–64	24
	Are we troubled by ambiguities in aggregation or by comparison of prices among countries?	65	24
V.	Differences between the recommendations in the Castles report and in this report	66–72	25
	Agreements	67	25
	Differences of opinion	68–69	25
	New recommendations	70	26
	Audit and evaluation	71–72	26

* This report was prepared by Mr. Jacob Ryten under the sponsorship of the World Bank, the International Monetary Fund and the United Nations.

VI.	Credibility and process: how the current process is seen	73–77	26
VII.	Special circumstances, critical views and dilemmas	78–83	28
	Special circumstances affecting ICP	78	28
	Criticisms one hears	79	29
	Dilemmas for future phases	80–83	31
VIII.	Possible answers and practical steps	84–96	35
	Possible answers	84–86	35
	Launching an ICP phase	87–93	36
	The role of a good handbook	94–96	40
IX.	More questions and answers	97–137	41
	The matter of GDP weights	97–100	41
	How much of GDP should be covered?	101–115	42
	The matter of GDP aggregation	116–120	45
	Bridge countries	121–123	47
	Proposal for a faster way of producing results	124–130	47
	A question of geography	131–137	48
X.	Steps to be taken in the short term	138–154	51
	Three necessary features	138	51
	A help desk	139–142	51
	Editing guidelines	143–147	52
	An analytical capability	148–154	53
XI.	Cost, recommendations and conclusions	155–161	56
	How much will it cost?	155–159	56
	Recommendations	160	58
	Conclusions	161	59

Executive summary

1. The following are the essential elements of the report.

If there were only one paragraph available into which to squeeze the essence of this report it would start by saying that there is reason to keep the International Comparison Programme (ICP) but that its results must become more credible and more useful. In order to bring about the required change, the United Nations Statistics Division should find a "world coordinator" with the experience and seniority that kind of a job must command. The coordinator should be asked to put together a financing consortium consisting of National Statistical Offices (NSOs) and international agencies to ensure the success of a next phase of ICP; rewrite the *Handbook of the International Comparison Programme*[a] to make it obvious that the process is open, objective and effective; and find a way of getting the results to the hands of users in a timely fashion. The alternative of not doing anything is the worst, but doing something credible will require additional expenses.

---

[a] Studies in Methods, Series F, No. 62 (United Nations publication, Sales No. E.92.XVII.12).

Why this report?

2. The purpose of this report is to examine the circumstances in which ICP finds itself 30 years after its creation and to make a judgement about its prospects of the next while. In order to do so the report addresses such questions as:

(a) Do the purchasing power parities (PPPs) and PPP-adjusted statistics produced by the Programme serve useful purposes, which could not be served by cheaper or better alternatives?

(b) Do the statistics produced by the Programme in their present form serve those same purposes with sufficient quality?

(c) If improvements are thought necessary, what kinds of measures should be taken to bring them about and why?

(d) How should those improvements be managed?

(e) How much might they cost?

(f) How might one go about securing the financing necessary to institute such measures?

(g) What are the first things to address if the recommendations included in the report appear to be reasonable?

3. Section II of the report addresses these questions and provides answers to them. It is chiefly designed for those whose interest is exclusively with the action-oriented part of the report or whose concerns with the Programme and its results extend no further than to be aware in general of what this report recommends. Subsequent sections address the questions one by one and expand on what led to the answers spelt out in the first section.

Appendix I* summarizes the results of interviews that the consultant conducted with users of the data, potential users and suppliers of basic information in national and international offices.

> The precise circumstances that brought about the report are recorded in the minutes of the twenty-ninth session of the Statistical Commission (New York, February 1997). It called for a consultancy to result in a report to the Commission on the state of affairs of the project to compare economic performance using purchasing power parity techniques. The terms of reference of the consultancy are spelt out in the annex to document E/CN.3/1997/3/Add.1. The direct sponsors of the report are the United Nations Statistics Division, the International Monetary Fund (IMF) and the World Bank.

4. This report is not the only one of its kind. Another consultant wrote a report similar in intent but its scope was the Organisation for Economic Cooperation and Development (OECD) segment of the Programme. That report has been in the public domain for the best part of a year and has been discussed at a number of official meetings.[1] Its assessments and advice are similar in some respects to those of this report but in others they differ. In appendix II,* there is a discussion in detail of the points of agreement and difference between the two reports.

5. Since this report is not designed as a thriller there is no point in keeping its chief conclusions to the very end. Those conclusions put very briefly are that ICP is a programme worth keeping but that its current condition, if little is done about it in terms of credibility, quality of output and survival prospects, is poor. A number of measures are recommended, some designed for immediate action and others to be spread over the next few years, but all conditioned by the likelihood of adequate financial assistance to the Programme. Another conclusion is that if little is done to rescue the Programme it will probably, like the old soldiers in the song, simply "fade away".

6. Much of what is said in this report is critical and may sound harsh to those who have worked very hard to promote the Programme's usefulness and uphold its integrity. They have performed a remarkable job under adverse circumstances and in an area fraught with practical and conceptual difficulties. The international statistical community should feel indebted to them. But it would serve little purpose if this report glossed over the shortcomings of the Programme and failed to show how they affect adversely its credibility. The sense of the report is to justify as closely as possible all the remedial measures it advocates, and that entails an unbiased view of the Programme's current situation and prospects.

---

* The appendices to this report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

[1] In particular at an OECD meeting on purchasing power parities, held in Paris in November 1997; and at the nineteenth session of the Working Group on International Statistical Programmes and Coordination of the Statistical Commission (New York, 10–12 February 1998).

Past criticisms and past reactions

7. The questions addressed in this report are not new. They have been raised again and again by users and suppliers of data alike and probably to the annoyance of the upholders of ICP, who have been struggling against increasing odds to maintain the integrity and the usefulness of the Programme. In a reply to a criticism voiced by Paul Samuelson about the World Bank's obstinacy in publishing "the wrong figures"[2] the then Deputy Vice President of the World Bank commented:

> "... the central tasks facing us [before starting to apply PPPs for the Bank's operational purposes] are expanding country coverage and making sure that ICP data are made available in a timely and regular fashion ..."

And his diagnostic was:

> "... they [many developing countries] have found the work financially burdensome and have seen little policy use for the resulting estimates. Second, some developing countries have feared that ICP results, which show higher gross domestic product (GDP) estimates ... may be used to the detriment of their standing in multilateral lending agencies. Third, the statistical capabilities of some of these countries have not readily supported a full-scale ICP survey".

The two steps designed to overcome some of the observed difficulties and mentioned specifically were "... that the excluded countries be covered through a 'limited commodity approach'" and "major multilateral financial organizations to cooperate with the Bank in a type of cost-sharing arrangement".

8. In addition to providing reasons why it agrees with the criticisms and the measures to deal with them, this report examines a number of additional critical views and advocates measures to deal with them. In particular, it deals at length with the Programme's lack of credibility in the eyes of both users and suppliers of data and examines ways of augmenting it. It looks into elements of quality, particularly its inadmissible lack of timeliness, and suggests ways of overcoming it.

9. Lastly, the report looks into issues of management and organization of the Programme, notes deficiencies in both and calls for a quick adoption of measures to strengthen the Programme's administration. It suggests what such measures might consist in and seeks to impart a sense of urgency in gaining and preserving momentum if the Programme is to be rescued. The key elements of its recommendations are somewhat similar to those used for the launching of the National Household Survey Capability Programme (NHSCP). They consist in:

(a) Finding a coordinator of stature and credibility commensurate with the importance that this Programme should have in the eyes of the international statistical community;

(b) Persuading as many advanced NSOs as possible to staff a handful of posts — at their cost — to support the coordinator. Those posts are to be used by their sponsors as mid-career traineeships.

10. This report is not about finding weaknesses in the current methods of imputation and aggregation of basic data. Rather, it considers that the advice featured in the System of

[2] Comparative tables of gross domestic product, converted at prevailing market exchange rates.

National Accounts, 1993 (1993 SNA)[3] should be implemented and that right now it is senseless to divert efforts to anything other than analysing systematically the differences in the aggregates reached by different methods.

11. Nor is this report about the design of a method that will help NSOs in their efforts at collecting comparable basic data. At this stage in the Programme's life and in the face of the problems that put its survival in the balance, neither more refined methods of aggregation nor improved sampling schemes are thought to be matters of first importance.

Are PPPs required?

12. There is a general answer to this question the appeal of which should be seen readily by the international statistical community. Twenty years ago the decision was taken to embark on what led to the most thorough and most expensive round of revisions to the United Nations System of National Accounts. The exercise had many purposes, but one of them was to confirm the existence of an international language for statisticians who compile macroeconomic statistics and ensure that they mean the same thing when they say it the same way. The investment that went together with the revision will only yield full returns when the data will allow us to compare both the rates of growth and the levels of the economy's broad aggregates. But in order to do so for GDP levels (and for the elements of GDP) we must not find ourselves deterred by the existence of different currencies or become exclusively dependent on the market exchange rates among them. Conceptually, theoretically and practically, the United Nations national accounts programme will only be complete when it encompasses ICP.

13. This view can be traced to an expert group meeting on ICP methodology and implementation and to the following reaction to its report at the twenty-seventh session of the Statistical Commission:

> "The Commission expressed support for the plan [that] ... there would be a departure from the costly practice of launching benchmark year comparisons every five years and a move towards integration with work on national accounts and consumer price indexes."

14. But general reasons of this kind are seldom found to be compelling by national Governments. For the PPP estimates to get increased legitimacy in the eyes of national government statisticians and above all in the eyes of their paymasters, they require describable and reasonably important policy applications. The interviews conducted as part of the process that led to this report produced convincing evidence that PPPs and PPP-adjusted GDP figures are indeed required for a number of policy analytical purposes. Indeed, were it not for perceived weaknesses in the current methods, important operational applications could have been found by now. There is more detail provided on these applications in the relevant section of this report. The following are four examples of important applications:

(a) A better assessment of poverty and its distribution, without which allocation of scarce funds to needy recipients may be less effective;

[3] United Nations, *System of National Accounts, 1993*, Statistical Papers, Series F, No. 2, Rev.4 (United Nations publication, Sales No. E.94.XVII.4); see sect. 16.103, items (a) and (b); this recommendation was implicitly accepted by the international statistical community and explicitly by the five organizations under whose aegis the 1993 SNA was published. For further details, see sect. IX below.

(b) Better founded judgements on IMF quotas and drawing rights for member countries;

(c) Better first guesses at setting exchange rates for those countries that are opening up their economy to foreign trade and investment;

(d) A clearer understanding of the effects of competitiveness on foreign trade, with its consequent effects on the quality of trade policy advice.

15. It goes without saying that PPP-adjusted GDP numbers should be an essential complement to GDP in constant prices, and that the two sets of estimates ought to play equal roles in explaining economic growth in a world of rapidly expanding international trade and investment. While none of these arguments may be found compelling, taken together they support the assertion that without PPPs, the system of international economic statistics required to inform policy-making, monitoring and evaluation would be considerably poorer.

16. None of the above answers questions about the usefulness of PPPs in their current condition. Those issues fall into two categories:

(a) Are PPPs with their current attributes of timeliness and reliability worth preserving?

(b) Are the current estimates of PPPs susceptible to improvement at a cost within the reach of the international statistical community?

17. The matters concerned with the results of PPPs and how they compare with alternatives are referred to in section IV below. In particular, there is a reference to what is known about the explanatory power of PPPs relative to the market exchange alternative.

Are ICP estimates credible?

18. They are not and therein lies most of the Programme's problem. "The value of intelligence depends on its breeding," says a well-known author of espionage novels.[4] And so it is with statistics. For viewed from the outside we can assess their potential value but not their reliability. To help us assess "breeding" we depend largely on the credibility of the process and on the confidence we place on those in charge. Most of this report discusses the "breeding" of ICP estimates. These estimates have been the subject of criticism and most of it is tied to the process of collecting, compiling and disseminating the statistics produced by the Programme.

19. To an extent unequalled by any other statistic in the international domain, PPPs depend on the intimate cooperation between NSOs and the statistical arms of international agencies. However, the mechanisms that support what should be both an easy and an intimate partnership are mostly embryonic. In some cases, they are simply not there. In the past, too much attention has been paid to how to aggregate basic data once those data become available, but insufficient attention has been given to how they should be collected in the first place. The steps advocated in this report are designed to strengthen the "breeding" of the data; to promote its value in the eyes of potential users; and to strengthen the vital links between national and international offices concerned with PPP compilation.

20. Naturally, the steps proposed cost money, and it would be foolish to pretend that the Programme could be tidied up free of cost. In fact, the report argues that the resources it consumes are small compared to the importance that the purchasing power parity statistics

[4] John Le Carré, in *Murder of Quality* (Victor Gollancz, 1962).

could acquire if they were only produced on a regular, systematic and timely fashion. There are ways of attempting to secure additional resources, and the report includes proposals on how to do it. There are of course alternatives — not doing anything and letting the Programme fade away or else deliberately hastening its demise. While the latter is an improbable course of action — the international statistical community is notoriously against capital punishment — the former is more likely and for that reason denounced in the report as an unfortunate option. This is essentially what the report says. From now on there is argument and detail.

Recommendations

21. The following are the recommendations of this report:

1. ICP should not be ended nor should it be allowed to languish. Indeed, if no adequate financial support can be found for it, languishing would be its worst fate.
2. Securing financing on a broader scale implies making a commitment to producing reliable and timely data, with well-documented methods and sound analytical commentary.
3. While the long-term objective remains that of estimating all the components of final demand, in a first stage the compilers should have the modesty of making do with price estimates for household expenditures.
4. The savings accruing from a more restricted price collection should be ploughed right back into the Programme.
5. The Programme must have a global or world coordinator.
6. The coordinator must be known, respected and with the demonstrated administrative and professional abilities (the expression "professional" involves a grasp of the complex of national accounting, economic applications, and basic statistics) to coordinate a project of this size and complexity.
7. A new phase of the project must start with a resolution adopted by the Statistical Commission. That resolution should follow the submission of a document that sets out unambiguously what is expected, why it is being done, what means will be used, what are the responsibilities and accountabilities of the participants, and what are the standards of quality that are aimed at in connection with ICP.
8. A Commission resolution should be the result and not the cause of mobilization of resources. The latter should be preparatory to final approval.
9. The next phase should be designed in such a way as to produce continuous information based either on a benchmark study or else updated through the help of consumer price indexes (CPIs) and exchange rates.

10. Additional resources to the project should be obtained by creating training positions attached to the coordinator and financed by sponsoring statistical offices.[5]

11. The dissemination activity must be shared between NSOs and the statistical arms of international agencies. Country statistical offices must be enlisted in order to give the project greater visibility and a stronger sense of relevance.

12. Interested parties (the United Nations, IMF, the World Bank, the Asian and Inter-American Development Banks, selected NSOs) should mobilize the required resources under the guidance of the world coordinator to:

(a) Prepare an ICP prospectus as a means for discussion with prospective contributors;

(b) Recruit the first set of short-term assistants after defining the modalities of their contract;

(c) Embark on the drafting of a revised and expanded handbook (probably manual is a better term) offering guidance to NSOs engaged in ICP.

Conclusions

22. No statistical programme with an international dimension needs central coordination and an effective relationship with NSOs more than ICP. The soundness of the Programme requires that both national and international offices play their role effectively. As a result it is more vulnerable than average to personality conflicts, small changes in budget, apparent lack of direction and so on. On the other hand, a strong hand, a feeling of commitment and purpose and the rallying of support on the part of NSOs can turn the situation around quickly. A programme evaluation is an opportunity to ask fundamental questions about the wisdom of keeping a programme alive. If the answer is "yes" and the cost can be made affordable, the other decisions to ensure survival should not be too difficult to take.

[5] The modalities of this proposal would have to be worked on and be subject to administrative and financial constraints imposed on the one hand by the United Nations and on the other by the sponsoring countries; but permanent stationing in New York, Washington, D.C., Paris or Luxembourg is no longer necessary as long as there is a good communications network that will allow for video-conferencing and intensive exchange of views via e-mail, phone and fax.

I. Introduction

This report is the result of an evaluation of the International Comparison Programme (ICP) conducted in the course of 1997–1998 at the request of the Statistical Commission and under the sponsorship of the United Nations, the International Monetary Fund and the World Bank. While the contents of the report were discussed with its sponsors, all opinions, recommendations and conclusions are strictly the author's.

23. ICP is a 30-year-old programme, tried and tested on several occasions. It is the origin of a large fund of valuable data successfully used as a complement to related data compiled in the framework of individual country's national accounts. There is a great deal of literature on the subject of PPPs, the estimation of which is at the root of ICP. That literature analyses the results obtained in the past, discusses alternative methods of estimation and aggregation, and shows conclusively why intercountry comparisons that rely on market exchange rates can yield misleading results for both policy design and economic hypothesis testing.

24. Nonetheless, ICP is in crisis, which is why an evaluation of its condition was requested. The crisis threatens the Programme's chances of surviving on a sound financial footing and seriously undermines the credibility of the numbers it estimates. Neither countries taking part in the ICP nor international organizations involved in its administration feel that the Programme is in equilibrium or that such an equilibrium is stable. This feeling which has been described as a "collective unease" has given rise to a few basic questions, which this introduction attempts to answer in summary form.

25. This report responds firmly to the question of whether it is worthwhile making an outright effort to keep ICP as an active programme with a resounding "yes". It finds that there is a sufficient number of serious applications of the results to justify such a convulsion. Those applications include:

(a) All intercountry comparisons that involve the level of economic performance;

(b) The rational allocations of scarce entitlements, such as IMF quotas or drawing rights;

(c) The determination of appropriate exchange rates for those countries that are in the process of opening their economy to international trade and investment;

(d) A better understanding of the factors that determine international competitiveness;

(e) A new light on the relationships between consumption, investment and economic growth.

26. Nonetheless the range of applications for which ICP results have been used does not include such compelling examples as the allocation of scarce funds assigned to the fight against poverty or the allocation of conditional credits for infrastructural investment under the aegis of the World Bank. If only these were among the purposes for which the programme is carried out, much of the uncertainty about its future would vanish, and so would the feeling of commitment and the willingness to take part on the part of NSOs.

27. Admittedly, the Programme is not in shape to justify adding the allocation of scarce funds to its objectives. Not only is its lack of timeliness deplorable but its results are not

generally accepted in the same way as such key statistics as the CPI or the gross national product of individual countries. However, the Programme's credibility and acceptance can only increase if all participants — NSOs and the statistical arms of international agencies — feel a profound sense of commitment to ICP's quality. And that sense of commitment will only come about if the Programme's applications constitute compelling reasons to support it.

28. There is a virtuous circle which has somehow escaped the reach of ICP managers. The fact that the Programme's objectives are not perceived as of key importance by NSOs or indeed by national economic authorities deprives it of an essential source of support. Without such support, the Programme's quality features — timeliness, reliability of results, transparency of methods — are seriously compromised. But the support will not be forthcoming unless the Programme's objectives are substantially upgraded and so on.

29. This report reflects on the causes of the ICP crisis, and recommends the taking of a number of steps designed to pull the Programme out of its current condition. The recommendations are built on the assumption that steps will be taken gradually, and that as the quality of ICP increases so will the amount of support given to it by individual NSOs. There are a number of initial steps though, without which the virtuous circle will continue to escape ICP's grasp. Such steps include:

(a) The appointment of a world coordinator to provide strong leadership — a sense of direction and a sense of trust;

(b) The immediate adoption of methods to improve ICP's timeliness, albeit on the basis of preliminary estimates;

(c) The systematic canvassing of potential participants in order to ensure that they understand ICP's scope, usefulness and potential importance to them (national authorities) of a well-managed programme;

(d) The clear demonstration of transparency in methods and applications, coupled with a willingness to share with participants the lessons of success and failure.

30. The international statistical community has not been known for its willingness to consign to oblivion programmes that have provided less than initially expected from them. This attitude is both understandable and prudent. And yet it entails serious risks. If strong support for ICP is not forthcoming and if the Programme is not given another chance to demonstrate that its results are generally useful and enlightening, the alternative of letting it languish is about the worst that could happen to it. This report should be viewed as a lengthy argument to prevent the worst outcome from happening.

II. A review of ICP

What is the subject of ICP?

31. At first blush one would think the matter raised by ICP is trivial. For a variety of reasons one seeks to compare the economic performance of one country with others. Insofar as individual countries adopt their own currency, comparisons require currency converters so as to express similar magnitudes in the same currency. The natural converter is the exchange rate. But exchange rates are volatile. They are partly dominated by expectations of how asset prices are likely to behave in the short term, and not all — indeed only a

minority — of goods and services produced in any one country enter international trade.[6] Accordingly, a more meaningful converter is proposed — one that is defined as the ratio between the number of units of country A's currency required to purchase in A the same amount of goods and services that one unit of currency of country B would buy in B.

32. Think of a very small universe in which there are two countries (Utopia and Ucronia) and two commodities (hamburgers and Coca-Cola). The currency in Utopia is the "bottle top" and the currency in Ucronia the "shaker". If one "bottletop" in Utopia purchases one bottle of Coca-Cola and one hamburger but it takes three "shakers" to make the same purchase in Ucronia, it follows that at an exchange rate of three "shakers" to one "bottletop" there is "purchasing power parity" between Utopia and Ucronia, whatever the market exchange rates between the two countries say or do.

33. All that is involved in this matter is that whenever Utopian levels of economic performance expressed in value terms are compared to Ucronian and until further notice (dictated by a change in relative prices), the converting rate of three to one is what should be applied. The use of this conversion ratio does not involve predicting what will happen to the market exchange rate of "bottle tops" for "shakers".

34. If in addition to comparisons, there is a need to compile information for the universe made up of Utopia and Ucronia, the rule to follow is to either express all magnitudes in "bottletops" after dividing those expressed in "shakers" by three, or the reverse, or converting both "bottletops" and "shakers" to a notional third currency, as long as the exchange rates preserve the proportions of three to one between our two hypothetical currencies.

35. If we lived in such a simple universe with a trivial number of commodities and constancy of tastes and technologies, that would be the end of the problem of purchasing power parities. But we do not. In the universe in which we live there are many countries, there are many goods and services, and tastes and technologies vary among countries and change over time. It is neither straightforward to decide what is the composition of the standard purchase — should it reflect the tastes of Utopia, Ucronia or some third country and if so which — nor to ensure that the goods and services selected both satisfy the property of being widely consumed in each of the countries compared and at the same time remain comparable from one country to the next.

36. It is also not straightforward to define how the results of economic performance expressed in different currencies should be added up. Various techniques have been proposed but they provide different results. And there is still no theory to support the unambiguous use of one form of aggregation over another, nor is there a body of analysis to explain the way in which the differences between different aggregations change over time. As a result, the question of how to compare the economic performance of countries using a common yardstick remains without a simple and convincing answer, although a great deal of work has been invested in improving our understanding of the problem and considerable intellectual ingenuity has been displayed in avoiding some damaging pitfalls.

37. Notwithstanding the work and the ingenuity that have gone into the theoretical literature on PPPs, serious questions remain about the validity, reliability and usefulness of the results. Those questions have found expression among users and potential users and at a different level among the national producers of the raw data required to carry out comparisons. It is those questions that have given rise to this report.

[6] Unless the country or area is an entrepôt, such as Hong Kong, China or Singapore, and even then many services do not enter trade.

How international comparisons enter into everyday discourse

38. International — and interregional — comparisons are very much part of everyday language. We say that it is more expensive living in New York than in Mobile, Alabama, or that Paris is less expensive a place to visit than Rome. We act accordingly. Institutions that have an international sweep seek to remunerate their employees in a way that compensates for different costs of living. Sometimes, those efforts are very systematic. For example, the United Nations has an elaborate method of ascertaining differences in living expenses through time-to-time and place-to-place surveys. Foreign ministries have similar concerns, and either borrow information from the United Nations or else conduct their own surveys. Statistics Canada conducts such surveys on behalf of the Canadian Government. The results are deemed to be necessary so as to ensure equity in the pay schemes of Canadian government officials working abroad. Many large firms with headquarters in Canada but with operations abroad consult Statistics Canada regularly in order to adjust their own remuneration packages. All these are examples of comparisons of consumer expenditures conducted in order to adjust incomes accordingly. Incidentally, similar efforts are made within large countries, where living conditions vary considerably in cost from one region to another in spite of having a single currency and no internal tariffs.

39. We also say commonly that Utopia is a much richer country than Ucronia, that A is poor but not quite so poor as B, and that C is at least four times as productive as D. Some of these statements are purely impressionistic. Others go further and require quantitative evidence to substantiate them. *Faute de mieux*, the evidence is provided by the application of market exchange rates. But the use of those rates breaks down easily in both space and time. In many instances, including some that are relevant to discussions on poverty, there are many market exchange rates and the most relevant are the ones that are frowned upon by the country's authorities. In other cases, the rate is artificially set because there is no free international trade or investment and the results of a conversion are likely to defy common sense.

40. We have experienced recently situations of very rapid decline in the fortunes of not one but several currencies, and we have evidence to suggest that the resulting impoverishment would be grossly overstated if one were to apply the new market rates without qualification. It is certain that one would not engage in time-to-time comparisons involving the production of goods and services using such rates.

41. Ultimately, the ICP results allow us to make such statements or should constitute a court of appeal whenever we have conflicting impressions. For these reasons, the discussion that they prompt should be not so much about their right of existence in principle but rather whether they are sufficiently well estimated for most of the objectives at hand. Whether or not they are more useful in their present condition than market exchange rates is also a legitimate question, but its answer depends very much on the policy decisions we seek to inform, the conceptual barriers left to overcome and the means we feel we should place at the disposal of those in charge of ICP.

Why was this report commissioned?

42. The state of affairs of ICP has been discussed repeatedly at sessions of the Statistical Commission. A brief outline of the Programme's history[7] is that it started out as a collaborative project between the University of Pennsylvania and the United Nations Statistical Office (as it was then). Six events or phases, the first of which started in 1968 and led to a path-breaking publication of results in 1975,[8] marked the project's life. Beginning with the third phase, the project was upgraded to a Programme, the University of Pennsylvania assumed an advisory role, as opposed to that of a co-director; and a five-yearly schedule for new benchmarks was adopted. By phase four, the project became regionalized and the central direction that marked its earlier life was abandoned in favour of central coordination. The only change since then is that the five-year schedule has been abandoned in fact if not explicitly.

43. Between the end of phase four and phase six, the programme has been marked by uneven regional performance, as well as by a substantial increase in the number of countries taking part in it. While European Union (EU) countries have virtually succeeded in integrating the programme with their regular activities and conducted it on an annual basis, in other parts of the world the activity has been spotty. In the case of non-EU member countries of OECD, their participation has not always been enthusiastic. Even though the OECD segment of the programme is part of a regular schedule, there have been questions about participation all along. One deterrent to a more stable state of affairs has been the disproportion between the resources set aside for the exercise and the magnitude of the tasks associated with it.

44. In addition to chronic financial difficulties, the project has had to face limited credibility on the part of a number of key providers of the data in terms of its conceptual foundations, its usefulness and the practical details regarding its execution. Because criticisms of this nature have not yet been fully answered, the tentativeness of the reactions fuelled increasing doubts about the value of the undertaking. These doubts were reflected first in the report of a small task force convened by the Secretariat to discuss critical problems in economic statistics. The report was submitted to the Statistical Commission at its twenty-ninth session at which the Commission:

> "Agreed on the need to conduct an evaluation of the global ICP to address the reservations by certain member States about ICP implementation and the uses of ICP results, and the need to seek ways to improve the credibility of ICP data. The Commission noted that the timing of an evaluation was appropriate ...".[9]

The Commission also suggested that the steering committee, in reviewing the terms of reference for the evaluation of the global ICP:

---

[7] See United Nations, *Handbook of the International Comparison Programme, Studies in Methods, Series F, No. 62* (United Nations publication, Sales No. E.92.XVII.12), annex 1, for a concise history of the Programme.

[8] I. Kravis, *A System of International Comparisons of Gross Product and Purchasing Power* (Baltimore and London, Johns Hopkins University Press, 1975).

[9] See *Official Records of the Economic and Social Council, 1997, Supplement No. 4* (E/1997/24), para. 13 (j).

"... take into account the very special problems involved in making comparisons between highly developed and developing countries, such as how to ensure that the items chosen reflect common characteristics and are representative of all countries being compared".[10]

45. The present report is not one of a kind but one of two. For purpose of breaking down the problem into more manageable components and because circumstances do not affect participating countries evenly, the ICP world was divided into OECD and non-OECD countries. The way in which ICP has worked in the OECD area is described in the "Castles report",[11] which was commissioned by OECD and discussed at a meeting on purchasing power parities convened by OECD in November 1997. The rest of the world is dealt with in the text that follows. Because the Castles report has been around for the best part of one year, there are references in this report to the conclusions and recommendations reached in it. There is also a discussion in appendix II* on how the recommendations for non-OECD countries differ from the conclusions and advice contained in the Castles report.

* The appendices to the present report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

[10] Ibid., para. 13 (c).

[11] Ian Castles, "Review of the OECD-Eurostat PPP Program" (Paris, OECD), document STD/PPP(97)5.

III. What this report is and what it is not

Existing literature and perceived problems

46. In spite of its apparent simplicity, the problems raised by the organization and management of a successful ICP benchmark are formidable. It is therefore understandable why the 1993 meeting of experts discussed at length alternatives to the burdens of a five-year benchmark. It would have been surprising if concrete alternatives had emerged, particularly alternatives that would allay the concerns about the Programme's usefulness and at the same time be considerably simpler than the current practice. Because of the Programme's origins and because of its chronic struggle to secure a sound financial footing, many alternative ways of proceeding have been looked at in the literature — not only the kind of literature produced in the context of the management of international organizations but also in academic circles.

47. Thus, there is an ample supply of articles on the properties of the different aggregation schemes used to date, as well as innovative proposals to adopt more sophisticated aggregator functions.[12] There is equally abundant literature on the biases incurred by alternative forms of aggregating and imputing. The Statistical Commission, at its twenty-fifth session, requested that a handbook on ICP be prepared on the grounds that there had to be an effective way of keeping the suppliers of basic data abreast of the theoretical progress that supported the programme. The *Handbook of the International Comparison Programme*[13] was indeed prepared, and includes clear discussion of the principal methods of imputing and aggregating basic data. It is not the purpose of this report to contribute to that discussion, largely because that is not the Programme's most serious vulnerability.

48. There are at least two features of ICP that appear to have been neglected or at best not given as much prominence as they should. One is the physical organization of the collection and editing of basic data, and the other is the form in which the results should be disseminated. The underlying assumption that explains the lack of guidance on editing must have been that international advice should concern itself with matters that are familiar to NSOs. The failure to think hard about the most effective way to disseminate the results is less understandable. Both matters are dealt with in some detail below.

49. Essentially, this report attempts to deal with a few key questions asked in the course of the 1998 meeting of the Working Group on International Statistical Programmes and Coordination of the Statistical Commission:

[12] See R. J. Hill, "Comparing price levels and living standards across the ESCAP countries using spanning trees and other aggregation methods" (Beijing, 1997).

[13] United Nations, *Handbook of the International Comparison Programme*, Studies in Methods, Series F, No. 62 (United Nations publication, Sales No. E.92.XVII, 12); while the *Handbook* is mostly designed to provide guidance to the suppliers of basic data, not one of them in the NSOs visited appeared to be aware of its existence.

Four basic questions[a]

1. Why is there a feeling of unease about ICP?
2. Why does the process adopted to estimate PPPs not inspire confidence?
3. Why is no notice taken of the results of ICP?
4. Why is there so much literature on aggregation and so little on basic estimation?

[a] These questions are paraphrases of points made by critics at various times and sessions of the Commission: question 1 was placed by Tim Hold; question 2 is from a comment by Bill McLellan; question 3 was asked by Ivan Fellegiat the twenty-sixth session of the Commission; and question 4 is based on an informal comment by Michael Ward. The author regards them as expressing most of the reservations that NSOs have concerning this Programme.

50. The first of the questions is sufficiently general to warrant saying that this report is an analysis of why there is unease. The second question is taken to address the fact that there is no documented process and no defined attribution of roles and responsibilities. The report devotes a couple of sections to proposals on how to make the process more systematic and more explicit. The third question is interpreted to show the dilemma created by poor dissemination policies. Either the problems that PPPs are expected to shed light on are unimportant in the eyes of the potential users (witness their lack of reaction) or else — if supply is to create its own demand — the results of the programme have been so poorly sold that potential users are not aware of their importance. The report describes what capacities ought to be created in the short run to test the proposition that PPP-adjusted statistics are required by real users to help solve real problems. The last of the questions concerns a longstanding failing in ICP. Most technical discussion is about how to aggregate elementary data, but comparatively little attention has been paid to the errors and pitfall involved in data collection. This may have to do with the fact that the participants in the debate are mostly applied economists and national accountants, whereas the effort of collection and compilation is in the hands of statisticians who have not taken a position on the quality and use of the ultimate results. The report attempts to establish a better balance by addressing in far greater detail issues concerned with "estimation".

IV. Why the "unease" with ICP?

Is ICP justified in view of its applications?

51. The applications of the PPP project warrant its existence at current levels of expenditure and indeed as argued later on in this report warrant a substantial increase in expenditure. The increase in expenditure should also be borne by participating countries, and this report suggests ways to enlist and organize their support.

Asking the question the wrong way round

There is a certain irony in the questions asked about the validity and usefulness of the ICP programme. For example, the ICP report for the ESCWA region shows that with one exception, the ranking of members is left substantially unchanged as we move from market exchange rates to PPP-adjusted categories of final demand. The question goes: if no significant analytical proposition is to be changed, why bother with PPP adjustments rather than using exchange rates after the removal of stochastic fluctuations by the use of sensible moving averages? After the First World War and the consequent disruption of the international exchange rate system, the question was put the other way round: should one bother looking at exchange rates if one had information on purchasing power? Thus, in order to determine what exchange rates should be adopted, the answer was to make judicious use of purchasing power comparisons and get to the right level as quickly as possible (see J. M. Keynes, *Tract on Monetary Reform*, chap. 3). The idea was that even if the market found eventually the right rate, the path that led to it could be politically intolerable. Accordingly, the prudent course of action was to make comparisons of PP and on that basis attempt a first fix of exchange rates, leaving the markets to eventually find the proper level. In the paper "Economic consequences of Mr. Churchill", Keynes shows the results of failure to consider relative PPPs in attempting to return to the pre-war level for the pound sterling.

Not surprisingly, after the collapse of the Soviet system, when the Republics of the former USSR and more particularly the countries in Central and Eastern Europe that had formed part of the bloc wished to fix exchange rates, they found that the existing calculations of purchasing power parities provided a necessary first approximation.

52. Among the applications found, the following appear to rank among the most serious:

(a) The possibility to make comparisons involving levels of expenditure across countries. It would be downright irresponsible to have invested the best part of the 1980s on the minutiae of the framework of an international system of national accounts to ensure that there was intercountry comparability only to deny ourselves the right to compare levels of economic performance where there is no common currency;

(b) The need to take into account elements other than those relating to market exchange rate when deciding on allocations of scarce funds. For example, no matter what administrative system is in place, can one afford not to take into account PPP-corrected economic variables when determining access to credit?

Are administrative and operational applications too serious?

This is a delicate issue. For example, the *Handbook of the International Cooperation Programme* is ambiguous on this matter and in the end offers neither guidance, nor does it put forward a way of progressing:

> "... At the world level in general, ICP results have not been used for assessment in the United Nations, or for concessionary loan rates in the World Bank. The principal reason for this is that both institutions had an operational system in place prior to the ICP, and there was natural reluctance to immediately change it when improved estimates became available. Further, the benchmark estimates were usually available with a lag of several years and only covered a portion of the countries of concern to the United Nations and the World Bank. It has been the position of the Statistical Commission at recent sessions that, at the world level, ICP results would not be used for administrative purposes ..."[a]

Why not? Surely the aim in producing a statistic is not to make it so pure that its objectivity is undeniable but at the same time its practical use is negligible. The reasons advanced in the *Handbook* are transient. The existing systems of assessment are likely to decay, and if ICP is to see its health improve it must produce results on a timely fashion. Is the matter of operational applications to be discussed at that stage? Obviously, more discussion is warranted. Notwithstanding the position taken by the World Bank[b] assuring countries that PPPs are not used for administrative purposes (officially correct), it is difficult to imagine how one would not take them into account at all even when faced with a need to specify a tie-breaker.

---

[a] See *Studies in Methods, Series F, No. 62* (United Nations publication, Sales No. E.92.XVII.12), p. 10.

[b] S. Ahmad, *"The International Comparison Programme (ICP): what is it and where does it stand now?"* (Beijing, 1997).

(c) Better understanding of how variables related to economic growth (GDP productivity, capital formation) interrelate. This is important not only as a matter of general understanding of long-term growth but also as a matter of interest to actual policy makers, particularly those in developing economies, who seek guidance from comparisons of their circumstances with those of similar countries or countries at the same stage of development;

(d) A variety of applications seeking to establish convertible exchange rates for countries that emerge from a system of inconvertibility and controlled foreign trade;

(e) A less distorted view of the extent of poverty and its correlates.

Testing the quality of PPPs: estimates of GDP per capita

A paper by Summers and Heston[a] written one year ago for an ICP seminar held at Beijing is the only one that asks (and answers) the following key question: "... Without doubt *correct* PPPs are to be preferred to exchange rates for conversions directed at output comparisons, but it is at least possible that available ICP estimates of the PPPs are of such poor quality that they are less accurate estimates of the correct PPPs than exchange rates." The authors quote from a well known study of comparative economic growth[b] which includes the results of regressions of GDP growth on a number of variables, alternately using exchange rates and available PPPs. The conclusion is that the "regressions strongly prefer the Summers-Heston version of the level of real GDP."

---

[a] R. Summers and A. Heston, *"Use of ICP results: a note on estimates of GDP per capita"*, Beijing, 1997.

[b] R. J. Barro and X. Sala-I-Martin, *Economic Growth* (McGraw-Hill, 1994).

53. If these are good examples of the serious applications of PPPs, why is the programme subject to so much doubt? For at least three reasons, spelled out below. The recommendation in this report is that all three should be dealt with.

Operational and research applications

None of these questions are asked within the EU, where (a) the comparison programme is integrated with regular pricing programmes; (b) there are well known operational applications that follow from the regular estimation of PPPs; and (c) the network of contacts and meetings to keep the programme relevant and up to date are both well established.

54. There is no continuous flow of information related to PPPs in the way there is in connection with exchange rates and with the CPI. The planning of the PPP as a worldwide exercise only once every so many years is the single most important factor that detracts from its importance in the eyes of users and producers.[14] Statistics that are of little operational consequence will not get the financial support that the difficulties of their production, maintenance and development require. Once they fail to get financial support, they will lose (if they have ever had them) the features required to be used for high-profile applications. Accordingly, the recommendations include urging the organizers of the Programme to consider producing a flow of continuous information on PPPs, rather than detailed news only once every five years.

---

[14] Probably one of the factors reducing the importance of certain statistical offices in the eyes of their Governments is that the most important figures that they are in charge of — the economic accounts of the nation — appears late, much too late to be of operational relevance.

55. The presentation of PPPs has been less than accessible, and the meaning of the findings of each new round has not been the object of dissemination comparable to, say, what accompanies the balance-of-payments figures or the CPI. It is difficult if one is a user to value a particular statistic if it is not obvious to see what story it tells and how important that story is. The report includes recommendations to improve matters in this regard, and gives examples of more interesting content. Progress in timeliness accompanied by an intelligent and attractive means of disseminating the information might get economic policy analysts better attuned to the importance of PPPs. As the results of the programme are featured today, they will ignore them.[15]

56. There is no well organized attempt to enlist the interest of potential users within countries, and in the third world the PPP exercise is to a large extent regarded as something of interest to international organizations but not to those parts of international organizations where decisions are taken. The report includes recommendations on how to enlist the interest of potential users in member countries.

Is ICP justified in view of the quality of data it is likely to produce?

57. The quality of the data provided is known and has been tested, albeit mostly in an academic environment. It has also been sharply criticized. The following may be the best way to characterize data quality:

(a) For at least two regions (Latin America and the region spanned by the Economic and Social Commission for Asia and the Pacific (ESCAP)), the results at the level of the major aggregates appear to be reasonable; furthermore, when tested (see Heston and Summers on ICP and GDP per capita), they suggest more acceptable behaviour than the same variables converted at market exchange rates;

(b) Notwithstanding this general finding, there are exceptions in both regions. The ranking of GDP per capita in the Indian subcontinent is questioned by experts in the area, and the results estimated for Mexico appear to be counter-intuitive because of where they place Mexico in the league table for the area;

(c) At lower levels of detail, the results are not generally interpretable, that is, there is a significant number of results that do not lend themselves to either ready understanding or else an acceptable economic interpretation of how they came about;

(d) A similar situation prevails in the OECD area, where the Castles report has pointed out three areas (aggregates of basic headings) where the results provided are open to serious challenge. The challenge in question is that the changes in the league tables from one benchmark year to another are not borne out by evolutions in either internal prices or exchange rates. The variations are too widespread and too radical to be explained by the vagaries of small changes.

In general, the combination of time and cross-section review of ICP results raises too many instances that appear counter-intuitive.

[15] When Statistics Canada announced the results of a bilateral comparison of purchasing power parities with the United States the news was received with profound lack of interest by the potentially interested ministries and by the press. The analytical presentation may have left something to be desired but by and large it was neither illiterate nor banal. The problem was that it referred to a situation three years earlier and therefore of little operational significance.

Is it better to have no PPP data than data produced by ICP?

58. Official data tend to be as good as we believe them to be and as useful as the use we give them. This is not quite true because there is an important time dimension missing to the statement. Assume that ICP goes on collecting data for many more years and phases. The accumulated data at the time they are used are more useful than if what are deemed to be data of indifferent quality today did not exist. What is missing from the data is an appreciation of how far off the mark they may be.

59. This observation should not be confused with some kind of a cardinal measure of reliability, for such does not and will not exist in the foreseeable future.[16] If we had such an appreciation, no matter how tentative in its initial stages, this would go a long way to establish credibility and credibility would be a sound basis for continued use. The measures of dispersion suggested in appendix III* take a first step towards indicating what possible errors there might be. But clearly this is an area where research would have a large payoff, as compared with more research into aggregation that may well be in a zone of rapidly diminishing returns.

What is the minimum investment required for a substantial improvement in the quality of PPP data?

60. There are two things that must be held separate: investment and the need to finance it. The investment cannot be negligible. The investment needed is mostly in two categories of people: analysts and process managers. The latter are required to coordinate, to put together the necessary financial package to launch new ICP phases, and most of all to convince potential but undecided users of the utility of the results and of the fact that if not intensively used they will not improve. For this last point to have a ring of plausibility about it, the results must be placed in the hands of users much more promptly than they are today.

61. The current expenditure — worldwide — on the programme is difficult to estimate, because there is no clear way of assessing the effort the NSOs have put into this Programme. The total cost (excluding OECD, Eurostat and EuroCost) is probably of the order of US$ 1.7 million for efforts outside the OECD area. These costs have been spread over a period of more than one year (probably closer to three), and include:

(a) Holding regional seminars;

(b) Diverting existing staff from their regular survey activities to ICP price collection;

(c) Conducting special surveys for such expenditures as non-residential construction, machinery equipment etc.;

(d) Data entry in NSOs;

(e) Travel to NSOs to provide support, guidelines and assistance;

(f) Processing, analysing and publishing.

---

* The appendices to the present report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

[16] Reviewers of statistical programmes with a background in auditing often ask for a numerical measure of compliance with standards and procedures and measurable error — the entire package to produce a single number expressing the "reliability" of the statistic. No such number exists and no one has found a way of computing a number for complex statistics such as the family of national accounts. This, however, does not preclude NSOs and statistical offices in international agencies from providing a qualitative appreciation of the series that they produce.

62. Of those costs, some $300,000 on average were spent by each of the regional coordinators and another $200,000 per region for the non-household expenditures, and the rest are the soft costs incurred by participating countries. The additional expenditures advocated in this report would probably add another $1 million for a cycle of benchmarks and updates, of which say half to two thirds should come from interested countries and the rest from a consortium of international organizations. The additions would be essentially for the salary and non-salary costs for one world coordinator and several regional coordinators, the offsetting reductions from scaling down or failing to estimate for the next non-household expenditures. Interested countries would finance the salary of the staff engaged in analytical work.

How would we know better-quality data if we saw them?

63. Would better quality mean lack of counter-intuitive surprises and better process? Better process can be defined and achieved. Indeed, that is what most of the positive recommendations in this report are about. But what about counter-intuitive surprises? One man's intuition is not like another. And yet we are all experts in international purchasing power. Coherence over time[17] would go a long way towards establishing the kind of faith that we have in the key national economic statistics. While in instances affecting the GDP or the CPI we may quibble at the margin over a particular rise or fall or whether the level is systematically understated, for practical purposes we take what the series show as given. No such automatic reaction applies to PPPs. The Castles report gives examples based on the publication of PPPs by distinct directorates in OECD.

64. For these properties — coherence and the absence of credible counter-intuitive surprises — to stick and for the method to have the openness required to discuss basic figures, transparency must increase greatly, and so must the capacity of international organizations to answer any challenge to its aggregates.

Are we troubled by ambiguities in aggregation or by comparisons of actual prices among countries?

65. There are doubts expressed about the two. The case in the Castles report had to do with the ranking of countries by either price or volume for selected basic headings. But in the course of interviewing producers of basic data who had taken part in one or two of the regional seminars held in order to launch phase six, such comments as the following were made with disturbing frequency:

(a) "... country X, which we happen to know well, must have interpreted the reporting guidelines incorrectly because all its prices are much lower than we know to be the case ...";

(b) "... countries x, y, and z produced unexplained numbers for rent and the coordinators did not feel they could intervene. The result is that for an important segment of expenditures we have distortion right at the level of basic data ...";

(c) "... the prices submitted by our neighbours cannot be right because they look as if they only apply to the capital city, which we know not to be representative and in any case cannot be compared with our prices etc. ...".

[17] The kind of incoherence over time noted in the Castles report is most damaging to the credibility of the PPP exercise. Until there is a plausible answer describing why instability of ranking over time happens in spite of the fact that no domestic or market exchange rate can explain what was observed, this damaging criticism will remain as a permanent feature of PPPs, much to the detriment of the Programme as a whole.

V. Differences between the recommendations in the Castles report and in this report

66. This matter is discussed in detail in appendix II.* The condensed version set out below is designed to highlight issues that are not covered in the Castles report and to mention the differences in points of view.

Agreements

67. Both the Castles report and this report agree on the notion that PPPs and PPP-adjusted macroeconomic variables have a very useful role to play, and that there is no serious alternative to their use for purposes of engaging in intercountry comparisons that involve levels rather than growth rates. Both reports agree on the need to integrate the ICP with national programmes in economic statistics, particularly with the national accounts and price programmes. Both reports agree that ICP has been underfunded relatively to the programme's objectives and that an infusion of resources is required to put the programme on a sounder footing. **These are by far the most fundamental aspects of the advice provided.**

Differences of opinion

68. The Castles report considers who should be responsible for the results of the PPP Programme, and advocates that it be exclusively the international agencies concerned with ICP. In this report, there is strong advocacy for the view that the responsibility should be shared. The grounds advanced are:

(a) Securing cooperation from NSOs is essential to the survival of the Programme. The chances of it happening increase if NSOs feel that they are accountable for the quality of an output that is as relevant to them as to international agencies;

(b) The credibility of the Programme is adversely affected if users and suppliers of data feel they are not privy to the methods and procedures that international organizations have used to aggregate the data. Explicitly shared responsibility forces greater transparency;

(c) The cooperation of NSOs is required both before and after the aggregation of data if the latter is viewed as an iterative process. The likelihood of that happening increases if NSOs feel as bound by the collective output as by the one specifically related to their country.

69. The Castles report advocates an about face in the conceptual target of the Programme. Specifically, it would see it concerned exclusively with household expenditures and its use limited to comparisons of real incomes. The grounds advanced are both practical (wise allocation of funds) and theoretical (the true purpose of the Programme does not involve the comparison of GDP and all the components of final demand). In this report, great sympathy is expressed for the practical arguments. If practical considerations of this nature are of consequence to OECD countries, a fortiori they must affect the rest of the world at least to an equal extent. But the present report considers that ultimately, the target of the

* The appendices to the present report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

Programme should be the adjustment of GDP, and it cites a number of examples — sufficient to be persuasive — of applications that require adjusting all the categories of final demand.

New recommendations

70. The Castles report is silent on the following matters no doubt because they are of less consequence for the OECD segment of the Programme. Even so, two of them play a role in striving to increase the Programme's credibility and should also be considered by the OECD's management of the PPP Programme. The matters in question are:

(a) Organizing the launch of a new phase mindful of the need to provide greater transparency, to define the role of NSOs and to ensure that there is proper coordination among regions and between regional coordinators and participating NSOs;

(b) Defining the job of a global coordinator, the support the job requires and what steps should be taken in the short run to improve the Programme's credibility and chances of survival;

(c) Improving substantially the timeliness of the Programme and examining how to minimize the impact of revisions whenever there is a new benchmark estimate;

(d) Outlining the kinds of analytical descriptions that should accompany the release of either benchmark or update figures in order to make the results of the Programme more interesting and relevant to users and potential users.

Audit and evaluation

71. One of the features that the Castles report and this report have in common is that both address the questions of whether ICP is worthwhile in the sense of meeting a specific demand and whether that demand is well met by the Programme as it currently exists. It would be better if (a) ICP itself had embodied provisions to be audited and evaluated on a regular basis; (b) both evaluation and audit were conducted by independent bodies; and (c) the reports were tabled at meetings of the Statistical Commission. This way, if circumstances were not favourable to the conducting of an audit — because there was no credible independent body or because the need to audit could be waived — the Commission would decide explicitly not to conduct it.

72. The sense of this recommendation is to add to the measures designed to promote transparency and build up the kind of credibility that the Programme lacks.

VI. Credibility and process: how the current process is seen

73. The way NSOs taking part in an ICP phase see the process should be of great concern to the Programme's organizers. The outline of the process is actually described in the *Handbook*, and is quoted below immediately after reporting how one of the country interviewees saw the events that marked phase six.

74. In outline form and starting from the moment when the head of the NSO took the decision that his country would take part in the forthcoming phase of ICP, a regional seminar was convened to explain the purpose, methods and procedures that would be used in phase six. The seminar was mostly devoted to:

(a) Explaining the methodology used for aggregation;

(b) Outlining the methodology to be used for price collection, including the specifics of:

(i) Price taking;

(ii) Data entry;

(c) The price survey was taken and data were submitted;

(d) Primary edits took place at the international agency in charge;

(e) There was feedback, but it was exclusively concerned with data entry problems;

(f) End of process (by default as there was no official notification to price collectors).

75. The process described above is no worse than the little that is said about the matter in the *Handbook*. Its only references (to how the process of data collection should be organized) consist in outlining the following steps.[18]

> "... Selecting and pricing representative items constitutes the most difficult, and typically quantitatively largest, most costly and time consuming part of ICP work for national statistical agencies and ICP organizers ... the main stages of this work are the same in all comparison. These are:
>
> "(a) Development of a list of representative items to be priced by a country; this list will be based on important items that are common to the existing national price data archives in the region or country group and draw on the core commodities;
>
> "(b) Collection of the price data not readily available from regular surveys;
>
> "(c) Submission of the national average prices for the items selected to the ICP organizers appropriate to the country;
>
> "(d) Checking the price ratios and parities at the basic heading level and correcting any unsatisfactory basic data."

76. Admittedly, the *Handbook* is very explicit in recognizing that the price collection stage of the Programme is the costliest and also the one that confronts NSOs with virtually all of the PPP operational problems they are likely to encounter. If ICP is only perceived for its nuisance value it is at this stage that efforts must be made to abate it, and if it presents a challenge it is equally at this stage that means must be made available to overcome it.

77. But there is an almost complete absence of references to who does what, how borderline cases will be dealt with, how a practical balance between characteristicity and comparability should be struck, and how to cope with unexpected events in the market place. There is no reference to what should be done if the existing CPI machinery only covers a small proportion of what is to be priced and at what point that may constitute grounds for going over the agreed list of goods and services once more.

[18] See United Nations, *Handbook of the International Comparison Programme* ..., p. 24.

Handbook of the International Comparison Programme[a] of the International Comparison Programme

The *Handbook of the International Comparison Programme* is mostly designed for NSOs, as it should be, and contains the kinds of subject one would expect such a handbook to contain. Thus, it has a section on purpose, concepts, articulation with SNA (which unfortunately was only available as a draft when the *Handbook* was put to bed), data collection, editing, classification, and the various ways in which the data are to be aggregated once the basic data are judged to be error free (or as free of errors as possible). At the end of the volume, which all in all has some 100 single-spaced pages, the *Handbook* has a useful annex on the history of the project.

In spite of touching on all the important issues, the *Handbook* is off balance in tone and in sense of proportion in what it selects as important. For example, the *Handbook* devotes no more than 11 per cent of total space to price data, mostly to definitional issues rather than practical issues of collection. From the point of view of an NSO wishing to join ICP for the first time or alternatively an NSO that has taken part in the past but is deciding on the margin whether to take part again, the interface — the point at which the international agency (ies) in charge interact with the NSOs is the critical issue. Assuming of course that there is a clear statement of use — internationally and nationally the *Handbook* is almost silent on this matter (see sect. VIII below for a detailed review of what other contents should be included in a future edition of the *Handbook*).

---

[a] Statistical Papers, Series F, No. 62 (United Nations publication, Sales No. E.92.XVII.12).

VII. Special circumstances, critical views and dilemmas

Special circumstances affecting ICP

78. ICP is a special programme that differs significantly from all the other data compilations in which international organizations are engaged. The following are the major differences between ICP and a standard programme, such as the data collection required to compile the United Nations *Yearbook of National Accounts*:

(a) The national accounts are estimated for domestic uses. Providing them to an international agency implies that at most they may be tweaked to conform to an international standard that is otherwise modified or disregarded. Even those countries for which direct domestic uses of such series as the national accounts are comparatively minor usually have indirect and international uses that are key to the Government's policies;

(b) PPPs are still in search of well-defined domestic uses, and in data producers' minds are exclusively required to support decisions and research projects undertaken by and for international agencies;

(c) Domestic series are compiled using domestic or international standards but are mostly used for domestic purposes. At the margin, there may be an international application;

(d) PPPs are compiled using very little of the domestic machinery developed essentially to support the CPI, or so it would seem. Only at the margin is the work for the CPI ever cannibalized by PPPs;

(e) By and large, NSOs know their domestic users. They interact with compilers in international agencies who in turn have their own users. Domestic users may or may not meet international users. For regular statistics, the contacts are at the level of the compilers (the OECD may be a partial exception to this model);

(f) In the case of ICP, the users as far as NSOs are concerned are the international compilers. The only role left for NSOs is to engage in the basic data collection, take part in editing if this is done collectively and take on a passive role in dissemination;

(g) For regular series, NSOs guarantee the quality of data collection and compilation — from the moment the raw information leaves the respondent to the point when it gets to the user's desk;

(h) In the case of PPPs, NSOs are in no position to guarantee anything other than the accuracy of their own basic data. But knowledge of all departures from comparability is essential to the success of the Programme;

(i) For regular series, the incentive for NSOs to produce their best estimate is clear and obvious. It is immaterial to country A if this objective is held high by country B or not. In the case of PPPs, the quality of A can only be assessed if taken together with that of B. If B does a poor job, the excellence of A's work is immaterial. This may be the **fundamental** difference;

(j) The most powerful promoter of improvement for A's regular statistics is public opinion as represented by users who look at A's series in a **domestic context**;

(k) In the case of PPPs, there can be no knowledgeable domestic critic, for the simple reason that it is only from an **international perspective** that the data can be critically appraised.

Criticisms one hears

79. The following are examples of the criticisms addressed to ICP and repeated by suppliers and potential users of the data:

(a) The PPP exercise is ultimately designed to meet the expectations and research needs of a few academic institutions. It does not justify the mobilization of a large number of NSOs and of analysts in a number of international organizations;

(b) The PPP project is imperfectly controlled at both country and at international level. Its quality constraints are not known, and the procedures required to carry out high level edits and imputations are neither documented nor replicable;

(c) The care and enthusiasm with which countries take part in the exercise is subject to great variability: to the enthusiasm of the person in charge, to the vagaries of the budgetary process in the country, to the amount of CPI machinery that can be mobilized for use in the PPPs and to the amount of international guidance given to those countries most in need of technical support;

(d) The results of PPP are counter-intuitive in terms of cross-country comparisons and not robust over time;

(e) International organizations have no coherent policy vis-à-vis the use of PPPs, and perpetuate the feeling of unease about the use of the results;

(f) There is no clarity about the policy use by international organizations on the results of ICP phases. For example, is it the case that when conditions for long-term lending are debated no participant in the discussion is supposed to refer implicitly or explicitly to PPP results?

(g) What is the use for policy and decision purposes of results that take on average three years to emerge?

(h) Different aggregator functions give different results. The differences in the results are not trivial. In some cases they are surprising, and we have not been able to explain them away in the same fashion as we can explain the differences between a national Paasche and Laspeyres index. One organization aggregates using the Geary-Khamis method and the country-product-dummy procedure, but another uses the Eltetö-Köves-Szulc method. Various indices are shown in OECD publications, but there has been little analysis of the differences between them in the last phases of ICP.[19] Hence we have a measuring apparatus (or several) but we are hesitant to relate the results to some established theory;

The Geary-Khamis and Eltetö-Köves-Szulc Methods, and the country-product-dummy procedure

These acronyms are explained at sufficient length in the *Handbook*. For the purposes of this report, it should suffice to say that the Geary-Khamis (G-K) method of aggregation revalues national expenditure categories at purchasing power parities, calculated as weighted arithmetic averages of prices prevailing in the region. The results are both transitive and additive. But it is also biased, in that it gives excessive importance to the prices of the more prosperous members of the set relatively to the less rich.

The Eltetö-Köves-Szulc (E-K-S) method of aggregation revalues expenditure components using geometric averages of all binary comparisons within the set (region), and makes the results transitive by a least squares procedure. While it avoids the G-K bias, it is not additive.

The country-product-dummy (CPD) is a multi-regression procedure that allows estimates to be made of missing country-products for a particular basic heading, by taking into account the maximum amount of relevant available country product information.[a]

[a] See I. Kravis, A. Heston and R. Summers, *International Comparisons of Real Product and Purchasing Power* (Baltimore and London, Johns Hopkins, 1978).

(i) Even if the problem of different aggregations were settled, the criticism goes that PPP can never apply to more than relatively small neighbourhoods (perhaps they should

[19] The poverty of analysis involved in the last three phases contrasts with the very thorough analysis of the third contained in I. Kravis, Al Heston, R. Summers, *International Comparisons of Real Product and Purchasing Power* (Baltimore and London, Johns Hopkins, 1978).

be no more than domestic spatial comparisons) where there is prima facie evidence that social habits, institutional structures, climatic factors, and consumer tastes are all roughly homogeneous, and where notions expressed with the same words mean approximately the same thing. The presumption that comparisons can be made at the world level is chimeric;[20]

(j) Even when applied to a roughly homogeneous region, the results obtained from ICP are not robust.[21] In particular, there is no coherence between time to time and place-to-place findings. For example: let us suppose that for a particular basic heading x, countries A and B at time *t* show A ranking above B in terms of say "real" consumption of the product. At time *t*-1, the ranking may be inverted even though internal evidence shows the two countries moving up at the same rate;

(k) The process that ought to lead from collection to aggregation of final results is iterative, complex and requires an enormous amount of coordination. Whether or not its coordination was recognized as necessary by the organizations responsible for the compilation is immaterial. The fact is that the actual control of the process is weak and poorly documented.

Dilemmas for future phases

80. In spite of a perceived indifference to the results of ICP, there is a significant difference between those instances where the PPPs are used for administrative purposes (EU) and those where they very explicitly are not. In the current condition of ICP, indifference is set against a background in which the majority of applications are research-oriented.

[20] See Arun Ghosh and Tarun Das, *International Comparisons of National Income: A New Methodology* (New Delhi, 1982).

[21] See Castles, op. cit., sect. 6.

The dilemma of "administrative purposes"

There is no doubt that the profile of an application affects enormously the status of a particular statistic. Thus the profile of the CPI both as a cost-of-living index and as an indicator of inflation has an importance second to none in the hierarchy of statistical office outputs. The census of population would attract far less support if its applications consisted exclusively in data for social and demographic studies and did not provide an accepted basis for transfers of funds, for electoral maps or for armed forces mobilization potential. Nonetheless, as soon as the administrative purposes of a statistic get to be widely known there is a pervading suspicion — however ill-founded that the compiled results reflect a bias presumably to favour the compiler or his paymaster.

In order to allay fears of misuse, particularly as PPPs provide one of the rare instances where no compiler is in control of the ultimate statistics, experts in charge go out of their way to make such statements as: "the statistics resulting from the PPP project are not used to affect World Bank lending conditions" or "The World Bank does not use ICP results to determine the terms of lending."[a] While understandable, this strong disclaimer has the effect of reducing the relevance of the basic information.

---

[a] See S. Ahmad, "The International Comparison Programme (ICP): what is it and where does it stand now?" (Beijing, 1997).

81. There is a need to review once again the balance between the perceived neutrality and impartiality of the results and their relevance to important decisions. If ultimately a project needs financial support and that support must come from Governments (or from budgetary authorities in international institutions), it is essential that applications relevant to government (and to the central mission of international institutions) be found. One of the first roles for the coordinators of the next phase is to consider this matter and make proposals for an enhanced use of ICP estimates, which should include, in however qualified a manner, the operational requirements of both international and regional organizations.

82. The matter of quality of the figures is one of the first into which future coordinators of the programme must look. In this respect, ICP is in a unique situation.

83. There is a need to devise a coordinated strategy for data collection, data publication and data analysis, the nature of which is difficult to specify in advance. Right now, there is no explicit approach that would enable the organizers of ICP to state where they would like to be in say five year's time. The fact that there is no framework classification of expenditures (because the Central Product Classification, for example, has not been adopted) makes it impossible to make such statements in advance. But one of the elements of more effective presentation of the Programme's objectives requires such an approach.

The paradox of quality

It is not — technically speaking — a paradox. But there is a key difference between how quality is treated in a purely domestic setting and in the context of PPPs. What constrains quality domestically is the budget. Otherwise, every marginal dollar can be used towards making an improvement in the target statistic. Not so in the case of PPPs. There would not be much sense if only country A invested in improving its quality, while all the comparators within the same region did otherwise. Indeed, country A is only interested in being as good as and no better or worse than the other countries with whom its regional total is about to be calculated. Unfortunately, there is no known way of making a detailed assessment of quality among countries, especially if some of the individual data have been gathered under the usual conditions of confidentiality.

Thus, for example, when Statistics Canada wanted to satisfy itself of the quality of the editing and imputation used by the OECD secretariat for the processing and aggregation of PPP data, access was barred on the legitimate grounds that for some member countries their submission was protected by confidentiality and the provision was only waived for the OECD secretariat. It follows that when the compiling agency requests improvements in quality, NSOs can only see it as worthwhile if they have a guarantee that the injunction is being complied with by all. However, there is no possible verification other than the integrity of the guarantor.

The paradox of the basic categories

The more one wishes to guarantee the quality of the results, the more specific the basic aggregation categories. ICP at the world level distinguishes 150 expenditure categories, and at the regional level (in particular for the developed regions) the product list has in excess of 1,000 goods and services. However, the greater the number of categories (basic headings or products to represent them) and the greater the detail of publication, the more open they leave the project to criticism resulting from visible inconsistencies among countries coupled with instability over time. In a situation where there is little credibility, every questionable result detracts more from what is already a reduced fund of good will. There is no alternative on this one but to pull oneself up by one's bootstraps.

Capitals, entire countries and the dilemma of geographical coverage

This is another example of trade-offs that have to be faced explicitly in future PPP rounds. It is very tempting to limit the scope of the comparisons to either the capital city of a country, or alternatively to its main business centre(s). The moment one starts to stray from these narrow confines, the number of complexities increases very rapidly. For not only does one encounter the traditional difficulties associated with intercountry comparisons but one compounds them by having to engage in indirect comparisons, involving comparing the capital with its hinterland as well as the two with their counterparts abroad.

Moreover, the chances are that in the capital city markets there is comparatively less imputation required, whereas the further one goes inland the more one finds situations that are either genuinely non-comparable or else require a great number of assumptions in order to be tractable.

On this matter, though, the *Handbook* is categorical (and should not be): "In contrast to the time-to-time price changes, where the necessity for absolutely identical items to be priced nationwide is not pressing, the ICPO makes more rigorous demands on comparability of specifications across observations within a country."[a]

If one shies away from full coverage one introduces distortions in the comparisons in a different way. In country A, the capital or the main commercial centre represent x per cent of the population, but in B it may represent some multiple of x. Accordingly, what is best — to compare A and B with different degrees of representativity in terms of share of population covered, or to only compare A and B once they have both reached a threshold in terms of population share? And do we mean population or income share? Is this comparison to be plutocratic or democratic? Is a plutocratic comparison what we want for such applications as "poverty"?

Suppose that whereas in country A a population or income threshold can be reached by surveying the capital city alone, in country B the desirable coverage can be reached only by going deep into the hinterland. This matter cannot be written off, but there is no policy other than what is advocated in the *Handbook*, which takes no notice of relative error. So this is yet another item to add to a much burdened research agenda.

---

[a] See United Nations *Handbook of the International Comparison Programme*, Studies in Methods, Series F, No. 62 (United Nations publications, Sales No. E.92 XVII.12), p. 34.

VIII. Possible answers and practical steps

Possible answers

84. The critical views listed in the previous section must be deflected and a proper resolution of the dilemmas must be worked out if ICP is to prosper. But the criticisms cannot be deflected all at once. In what follows, in particular in section X below, there is an outline of possible measures, including steps to be taken in the short run, such as:

(a) A clear statement — well beyond what is included in the *Handbook* — must be drafted about the uses of PPP results in addition to those that are more research-oriented in nature. A good starting point would be to recapitulate the uses specified in the study by Kravis et al,[22] and to reflect on how those uses have helped our understanding of a range of economic issues;

(b) A supplement to a statement must be appended drawn from such issues as allocation of funds to poverty (see paradox of "administrative uses" covered above and choose language clearly);

(c) The process underlying the new phase (see paras. 87–93 below) must be outlined, together with dates and times and financial means secured;

(d) The objectives of the research that is being carried out in parallel with the new phase should be stated, with an explanation of how they will contribute to an improvement in the quality of the information produced by ICP. The following are examples of lines of research that are required:

(i) Economic analysis of the differences in results obtained by use of different aggregator formulae, to reassure those who see as a major obstacle to the project's usefulness the fact that different aggregation schemes produce seemingly very different results;[23]

(ii) Further analysis of the problem of linking paths in geographical chain indexes;[24]

(iii) Finding defensible ways of combining cross section with time series analysis in order to get rid of the instability of ranking at the basic category level.

85. The ICP organizers will have to:

(a) Persuade country suppliers of the importance of the project;

(b) Find country users prepared to provide the necessary support for the project at country level;

(c) Ensure that the standards of the exercise, its objectives and techniques are well understood, and most important, that they are applied consistently (applied consistently does not mean applied identically — it means that the equilibrium between characteristicity and comparability, which is probably country specific, is respected throughout);

(d) Develop an ongoing feedback mechanism so that over time there is a better balance at the country level between characteristicity and comparability.

---

[22] Op. cit.

[23] If only care were taken in analysing the differences in G-K and E-K-S in concrete cases, what seems like unbridgeable differences would come down in size very rapidly.

[24] B. Szulc, *Criterion for Adequate Linking Paths in Chain Indices, in Improving the Quality of Price Indices*, Florence, 1995.

86. Above all, the ICP organizers will have to have a far more systematic approach to the organization of a phase.

Launching an ICP phase

87. Actually, there are two purposes served by this section of the report. First, it gets into detail because the material available to date has been almost exclusively concerned with conceptual and technical matters and has literally paid no attention to the details that confront NSOs when they take part in one of the ICP phases. Second, the purpose of outlining a detailed structure for the process to be adopted is to meet some of the sharper criticisms of the Programme[25] levelled at its lack of organizational structure.

88. In outline form, these are the principal stages to go through in launching a new phase of the ICP.

Setting the stage for a new phase

89. The following steps are to be taken at the global level:

(a) Statement of problem (assuming that in spite of scheduling a new benchmark every five years the organizers of ICP are asked to justify each phase individually; to do so, they would come before the Statistical Commission and announce the intention to carry out a new phase, make reference to its main challenges, and provide target dates for publication, rough cost estimates and so on):

(i) This step has to be very visible. It should give rise to an explicit decision by the Statistical Commission and should be handled in two stages. The first would endorse the desirability of a new phase and issue an instruction to the regional commissions to take up the proposal with their members at regional directors' conferences;

(ii) At a subsequent meeting of the Commission, members would take note of the reactions recorded in regional meetings and decide whether to proceed with the announcement that a new phase was to take place — and if so with which participation (regional conferences of directors would collect bona fide intentions to take part), with what financial arrangements, under whom as overall coordinator, and with what kind of a commitment for publication of the results;

(b) Estimation of costs and identification of sponsors. In parallel with the staff work required to consult, collect statements of intent, calculate the workload, inform the Commission and so on, staff work would also take place to provide realistic estimates of the following:

(i) Overall resource requirements, broken down into:

- NSO activities ranging from collection to data capture;
- bilateral technical assistance;
- multilateral meetings;
- data processing of the results;
- analysis of results and feedback;

[25] These are the ones made at several sessions of the Commission, and relate to the general feeling of unease that surrounds the programme; the absence of a coherent process; and the lack of reaction on the part of potential users to ICP data releases.

(ii) Establish the capacity of participating NSOs to finance domestic activities and take part in multilateral events;

(iii) Assistance in kind from NSOs interested in sponsoring the programme (see section X below for a proposal to secure such assistance);

(iv) Requirements from a consortium of interested international and supranational agencies;

(c) Appeal to participants by outlining the extent to which their cooperation is necessary to ensure the success of the programme:

(i) Immediately after the preparation of the financial estimates, it is incumbent to prepare a prospectus on what is of interest and should be known about the forthcoming phase of ICP: its scope, objectives, cost, technicalities, direction, date of completion and so on, in addition to the benefits that the information compiled is certain to bestow. The prospectus will constitute the basis for discussion of what financial help or help in kind is required to increase the chances of the project's success.

(ii) The actual announcement — the second step in the process of consultation with the Statistical Commission — should include the results of the appeal.

The launch

90. The steps listed below are to be taken at regional level.

(a) Identification of national interlocutors. These will tend to be those responsible for national accounts, the CPI and possibly other price surveys in NSOs;

(b) Convening of a regional seminar for the national experts to meet their regional ICP organizers and discuss the modalities of the new phase. The following is an outline for the agenda of the seminar:

(i) Define objectives of seminar and how they relate to the objectives of the new phase;

(ii) Discuss the constraints under which operations are to take place, the communications available for consultation and the mechanisms for settling differences of opinion in difficult cases;

(iii) Agreement on the common list of goods and services and corresponding descriptions;

(iv) Agreement on other aspects of price collection: sampling outlets, dealing with discounts, treatment of rents etc.;

(v) Agreement on transmission modalities: establishing links and defining procedures for the use of a regional help desk;

(vi) Agreement on a timetable for collection, data entry, submission etc.;

(vii) Agreement on procedures for editing and imputation;

(viii) Discussion of aggregation;

(ix) Assignment of responsibilities;

(c) While it is not necessary that all regional seminars be synchronized, it is useful and indeed essential that the proceedings and agreements reached in one be documented and made available to all the others.

The collection phase

91. This step takes place at NSOs but cannot proceed successfully unless there are frequent contacts between the regional coordinators and their national counterparts, bilaterally and multilaterally. The stages to consider are:

(a) Collection (institution of new price surveys, where necessary), indirect measurement, verification, settlement of doubts and documentation;

(b) Transmittal of data, together with documentation on conventions adopted, assumptions, difficult cases, exceptions to agreed procedures etc.;

(c) Feedback to regional coordinator and other NSOs in the region.

Seminar to review transmittal

92. Unless there are serious problems that are detected as a result of surprises at the aggregation stage or from the fact that the results are seriously questioned by any one of the agencies taking part in the exercise, this seminar would be the last stage requiring direct NSO involvement. The discussion would take place around a display of what price distributions for the region look like. Appendix III* presents a proposed design of tables that would constitute the basic documents for the discussion on editing and imputation. The agenda for the seminar should include discussion of the following points:

(a) An examination of all those cases where the dispersion of prices is an indication that there are inconsistencies in the data collection, in the sense that different agencies collected different things. The conclusions that might be drawn from such an examination are that: the agreed specifications were insufficient; no proper use was made of the help desk; the help desk did not work in this instance; the heading is much too broad and does not lend itself to tighter specification; or the initial agreement on what the specifications should be was inadequate and has to be reworked;

(b) An examination of those commodities where dispersion is unacceptably high but happen to belong to the "core".[26] For those cases, an immediate decision will have to be taken, and if so it should be a consensus decision in which both NSOs of member countries and the coordinators are involved;

(c) An examination of the change in rankings from the last phase. For each major change, there ought to be a satisfactory analytical explanation provided, combining internal price changes for the commodity or commodities affected, changes in exchange rates vis-à-vis some foreign supplier, changes in weights etc. In fact, one of the most damaging criticisms levelled against the Programme stems from the fact that it does not provide convincing explanations for cases — which at first sight appear counter-intuitive (see the Castles report for examples of counter-intuitive changes in ranking);

(d) An examination of weights as derived from the national accounts of member countries. For reasons which are understandable, gatherings to discuss PPPs turn out in most cases to be discussions of prices. This is partly because of the background of the participants but also because the data on prices are much more in the domain of everyone's daily experience and invite general comment. This is a carryover from discussions of the CPI,

* The appendices to the present report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

[26] The idea of core commodities — commodities which are common to all regions and must therefore have the highest priority for pricing purposes — is defined in United Nations, *Handbook of the International Comparison Programme* ..., p. 120.

in which weights are fixed and in any case series are extraordinarily robust to structural change. The same is not true for PPPs or rankings of GDP components. Small changes in relative weights can bring about significant changes in the position of any one country within its region.

(e) The agenda for a three-day meeting (this is an arbitrary notion, and the more that can be done ahead of the meeting the shorter the time required; however, there is enormous virtue in having a tête-à-tête encounter, particularly to discuss changes to initial agreements) would therefore run along the following lines:

(i) Presentation of first aggregation of GDP and examination of principal changes since the last round. Discussion of important changes;

(ii) Discussion of core basic headings (from the point of view of prices) which require revision or explanation;

(iii) Discussion of the dispersion of weights in questionable cases;

(iv) Decisions on matters that require immediate action and specification of work to be done upon return to respective NSOs;

(v) Agreement on a cut-off date.

Aggregation and feedback

93. This step is strictly in the hands of the regional coordinator and the ICP regional organizers, as follows:

(a) Complete aggregation;

(b) Convene seminar if there are residual difficulties that need some form of new collective agreement;

(c) Agree on conclusions: ranking, regional GDP, country shares in regional total and its components etc.;

(d) Feedback to NSOs;

(e) NSO sign-off.

What after regional totals are estimated?

After completion, it will be the task of the regional coordinator to prepare a press release with the results of the latest phase for the region. Preparation will require references to two sets of data: how the current phase differs from the previous and how the region has fared relatively to other regions. It is not realistic to create a global calendar with all regions synchronized on their release date. Accordingly, cross-section comparisons may require the use of preliminary data for other regions.

It is important to feature prominently how the current estimates differ from the previous ones in at least two senses: how the region has fared relatively to the rest of the world, and whether the winners and losers on the previous occasion were pretty much the same as currently or whether there were dramatic changes of fortune.

The role of a good handbook

94. Having a good handbook is not a sufficient condition for restoring the fortunes of ICP but it is a necessary step towards stabilizing and eventually improving the Programme. The elements listed below are primarily designed to provide more guidance to NSOs than what is provided by the current *Handbook*; to increase the degree of transparency of the Programme; and to ensure that it shares the international statistical infrastructure, without which it cannot be properly integrated with related statistical outputs.

95. The availability of a comprehensive handbook is an integral part of the improvements that ICP requires. A comprehensive revised edition of the *Handbook* should include the following elements:

(a) An explanation of the rules for editing and imputing applied by the international agencies in charge of the aggregation of basic data;

(b) A description of how those editing rules will be applied, including the notion of a collective review of edit failures, a step-by-step improvement in the quality of the descriptions attached to the standard list of goods and services priced etc.;

(c) A description of the services provided by a help desk, with particular reference to advice on departures from agreed descriptions or the need to submit replacements where goods and services that were present in the market no longer exist;

(d) A description of how the PPPs and adjusted GDP series will be disseminated, accompanied by examples of the type of analysis that should be carried out once the estimates are calculated. Insofar as the analysis is concerned, an accompanying description should be provided of how NSOs will find out about proposed analytical texts so that they can emit reservations about its nature;

(e) A stricter set of statements about the pricing rules for ICP and how they agree with the standards laid out in the 1993 SNA;

(f) A concordance (in the strict sense of the term) between the list of basic headings and CPC, Version 1.0, via the Classification of Individual Consumption by Purpose (COICOP);

(g) A section dealing with features that should be incorporated in household expenditure surveys so that broader international agreement can be secured. For example, all countries should agree to abide by COICOP as the outgoing classification. The classification used for data collection can be quite different though, as long as it is properly related to the outgoing classification;

(h) A description of the steps that will be taken in order to accelerate the compilation of new PPPs so that users in government do not regard them as data of archival interest.

96. In addition to these features — which are no more than a selection of items that would be of considerable help to NSOs — the strategy of publication of the revised *Handbook* ought to be rethought. For example, it might be worth investing in a CD-ROM that combines a data entry system, such as MOSAIC, with the *Handbook*.

IX. More questions and answers

The matter of GDP weights

97. The matter of GDP-derived expenditure weights has not received the attention it deserves in either the regional seminars held in the course of ICP phases or the *Handbook*. There is an interesting parallel between this relative silence and the treatment of expenditure weights in the CPI. Whereas there is considerable discussion of functional forms and sampling survey designs for price collection, there is relatively little on the systematic errors in weight determination that come about because of the difficulties in conducting household expenditure surveys. And yet the latter are arguably the most difficult surveys any NSO is likely to conduct.

98. There are several issues that deserve special study:

(a) What is the distribution of household family expenditure age within a region? If within a region there is rapid change in real income but considerable variation in the age of the last family expenditure survey, there will be a new bias in the determination of regional expenditure weights for those categories that are likely to have evolved fastest.

(b) What are the classifications of expenditures used in the household surveys, and how important are intercountry inconsistencies? Inconsistencies of this kind are usually carried over into the estimation of national accounting structures. Consider that they are inconsistencies across countries and not within countries.

(c) Do we have the right balance of detail in the categories for which we require expenditure weights, or — as is the case with characteristicity and comparability — do we need a new examination, taking into account known errors and biases?

(d) What can we learn from the distribution of regional weights within any one expenditure category? Should it be considered individually or amalgamated with other categories?

99. The comments above apply to the other categories of final demand, except that for capital investment the method of data collection induces far greater comparability, and in the case of the non-marketed portion of public sector activity other errors are more likely to swamp the results.

100. This report offers no answer to the questions above but suggests that they be given an important billing on the ICP research agenda.

Errors in expenditure weights

In referring to the expenditure weights required for PPP estimation, the *Handbook* refers to them simply as "data needed" but does not venture into speculation about the precautions required to ensure comparability or the consequences of likely error. But this is a case where it matters not only what biases there may be in the national data but also whether in taking the data for the various members of a region there is any likelihood of introducing new biases.

Consider the following possibility. In the case of most countries taking part in ICP, detailed expenditure weights will be derived variously from a household expenditure survey, an economic census and industrial surveys. The data will have been "homogenized" through some form of commodity flow analysis and brought to a point where they are internally consistent. Assume that there is a basic heading that is of importance for an area as a whole but that its importance within the area varies considerably from country to country.

Consider that, other things being equal, if a component is not important the resources available to improve the quality of its estimates will reflect the modesty of its contribution to the total. And consider as well that the most typical error of all is in coverage, and that it affects disproportionately those sectors that are relatively new; those that do not show a trend towards concentration but continue dominated by the many small enterprises that make it up; and those that have no special obligation to register their activity because it is not regulated.

Taking all these considerations into account, it follows that if country A has as one of its small expenditure categories a but its neighbour B has it as a large category, *caeteris paribus* the total a in A+B will be biased towards B's share. It follows that in calculating average prices using the G-K method, expenditure on a in B will weight the area average even more heavily than it should.

There are a great many assumptions in this speculation. But its point is to show that in cross-sectional comparisons we can no longer invoke the alibi of constant error in order to justify the reliability of the rates of change. There is another point and it is to encourage research on the possible effects of a particular distribution of errors within an area on that area's total and on the PPP-adjusted shares within the area.

How much of GDP should be covered?

101. The issue of GDP coverage is one on which there are differences between the Castles report and this report. The concerns in attempting to reach a considered view on the matter fall into at least four categories:

(a) Conceptual (vis-à-vis users and potential users);

(b) Strategic (vis-à-vis both patrons and users);

(c) Tactical (vis-à-vis possibly disappointed users);

(d) Operational (vis-à-vis patrons and NSOs).

102. The criticism that combines most of these matters runs roughly as follows. The difficulty with which the different components of final demand can be assessed varies a great deal. Thus, nothing is seemingly more difficult than to estimate PPP-adjusted consumption of goods and services by the public sector. In the Castles report, these expenditures are referred to as "comparison resistant". Nothing is more costly than the PPP adjustment of expenditure on capital formation, particularly non-residential capital formation. Indeed, the solutions adopted for the latter cannot avoid the engagement at high cost of specialized advisers to ensure that the technical specifications of standard models are being adhered to.

103. It is questionable, in the light of the applications of PPP, whether full adjustment of GDP is required. Rather, such applications as the policy on poverty or allocations of either grants or borrowing rights appear more closely related to adjusted national income for which the proper component on the expenditure side is the expenditure of consumers as the proportion of income that is not spent and can be rightly thought of as a bundle of consumer goods and services forgone. The merit of concentrating on consumer expenditures is that it focuses on consumer prices and can lean heavily on existing machinery — knowledge, a panel of retail outlets, experienced price takers and so on.

104. Lastly, it is not tenable to argue that comparisons of productivity are just as important as the previous applications. To make it stick it would be necessary to adjust GDP by industry product, and that has been rejected long ago on grounds of complexity and expense. It follows that the right policy is to abandon all attempts to compare non-market sectors or alternatively sectors the complexity of which is such that any attempt to deal with them is bound to unbalance the budget. Not only are they difficult to deal with but it turns out they are not even necessary.

105. The Castles report includes a very articulate argument developed along these lines, and ends up by recommending that as a matter of concept, strategy and management of the current situation the calculation of the non-household components of GDP be stopped. That argument was met by the OECD secretariat with the following counter-arguments:

> "... The [Castles] report recommends that the deflation [PPP adjustment] of non-market services by employee compensation be stopped ... Generally countries preferred that the OECD and Eurostat should work on improving the representativity of the wage data collected in the short term etc.", and "[the Castles report] contends that, besides being of questionable reliability, the PPPs for gross fixed capital formation have little analytical use made of them. Most countries were of the opinion that economists, researchers and others have shown considerable analytical interest in these PPPs and that, rather than abandon them, the OECD and Eurostat should work on improving their accuracy."

106. This is insufficient reason, particularly for the non-OECD area. First, if it is the case that the Programme is facing a crisis of credibility, questions cannot be answered in terms of "countries felt that", particularly as it is precisely national users and potential users who have shown very little interest in the results of the Programme. The argument would have to run along the lines that if less than all components of final demand were estimated potential users would not place their faith in the programme's results. But that is far from proven. Second, the question that has to be put to country delegates who make no decisions on the financial implications of their advice and are not involved in the budgetary allocations to international statistical agencies is whether the course of action advocated in the Castles report is the wisest given existing and expected financial constraints. Lastly, questions about priorities must be focused very sharply. There are users for all kinds of information, but

users dislike having to choose between alternatives. For this reason, it is the role of NSOs and international agencies to show alternatives and in the end make their own assessment about the wisest form of allocating their resources.

107. On grounds of efficiency, there is no question that attempting to do everything at the same time produces more questionable results than otherwise. But if the most important target is relevance, the concept should be geared to what is most relevant. If in the end we want to say something about growth, what matters is the relative evolution of GDP. The question to put is whether by wanting to be fully relevant we run the risk of not achieving any significant result.

108. One other question is whether the approach adopted for OECD countries or, more narrowly defined, for the EU countries needs to be the same as the approach adopted for the rest of the world. If the answer is "yes", whatever scale of difficulties is found for OECD will also be found for rest of the world countries, not all of which are fortunate to have such well-organized NSOs as the former. If the answer is "no", in this particular instance we are forgoing *ab initio* the possibility of estimating expenditure on a world GDP in which all constituents are PPP adjusted.

109. This is an instance where long-term and short-term objectives may not coincide. In the long term, we should aim at estimating GDP and its expenditure components because applications will require some underpinning of changes in GDP. This does not prejudge the issue of whether one should attempt to estimate a GDP for the world as a whole. Nor does it imply that because we should aim at doing this over the long term we ought to do so immediately.

110. We would have to prove the following:

> Assume there is a constant budget that can be allocated either to consumer expenditure as a whole or else to the various components of final demand. If we allocate it entirely to the former, quality is significantly improved. If we allocate it to the various components of GDP, quality remains at its current level.

111. Since most users will want a statistic for GDP per capita, in the absence of an adjustment of each component of final demand they will estimate it by assuming that the adjustment factor for consumers' expenditure can be successfully generalized to other components. Assume further that consumers' expenditure is some proportion α of GDP for each of the countries in a given region. Assume that each component has a share in overall ϵ (error) which is proportional to its share in GDP. Assume that we seek to minimize overall error.

112. Let us now assume that we concentrate on a better estimate of consumers' expenditure to the exclusion of all else, as a result of which we succeed in halving its error. But in exchange, we estimate the rest of final demand by extrapolation, as a result of which we double its error of estimate. Our overall error becomes:

$$1/2\alpha\epsilon + 2\,(1 - \alpha)\,\epsilon$$

or

$$1 + \epsilon\,(2 - 1.5\alpha)$$

113. Simple algebra suggests that for the overall error to increase it is necessary for the proportion of consumers' expenditure in total GDP to be less than two thirds. The upshot of the exercise is to show that while it is very probable that on average in third world countries the proportion of consumption exceeds two thirds of total GDP, it is not likely

that the relationships between consumers' goods and services and the other components of final demand are such that extrapolation doubles the error.

114. The operational reasons against attempting to do too much in difficult domains is well spelled out in the report of the Secretariat of the Economic and Social Commission for Asia and the Pacific (ESCAP) on the 1993 ICP results. While these are couched in mild language, one recognizes between the lines the concern and possible frustration with the exercise:

> "The standard specifications of machinery and equipment are very different from those observed in the reporting countries; the machinery and equipment on the list are not available in the reporting country; machinery and equipment imported from different countries, though of similar capacity in terms of performance, differed enormously in price. ... Commodities were found either to be not available or obsolete for reporting countries ..." and "... While it was difficult for the reporting countries to collect prices on machinery and equipment, it was even more problematic to make price adjustments for quality differences, in an attempt to ensure comparability."

115. The recommendation is the following:

(a) Countries outside the OECD in general should tackle the matter of consumer expenditure first and foremost, rather than dividing their resources into two compartments. Conclusions about GDP as a whole could be derived either through extrapolation or alternatively by projecting existing measurements — to be treated as benchmarks;

(b) There should be an explicit target for institution of proper estimates for non-residential construction and eventually for complete coverage of GDP. Assume that there is a plan that calls for a world benchmark[27] once every five years. The target might be to complete the components of final demand over a period of 10 years, with progressive introduction of new categories. The benefits would be twofold: better focus for overall expenditures and the possibility of learning from the successes (and the failures) of EU country attempts at covering the entire range of GDP.

The matter of GDP aggregation

116. A matter that is mentioned often as one of the examples of the conceptual fragility of ICP is the possibility to reach different results for such aggregates as regional output, depending on which aggregation function is adopted. There are two major contenders (although there are more; the spanning tree method, which is one of the more promising developments, has not yet reached the stage at which any of the international organizations concerned with the publication of the Programme results is ready to consider it as an alternative aggregation method).

117. The matter of different approaches to aggregation emerged at several of the interviews as a criticism of the Programme in any one of at least three guises:

(a) Because there were alternative methods of aggregating and as a result there was no single official figure;

(b) Because the existence of alternatives confirmed the impression that the entire Programme was at an experimental stage, and should be treated as no better than a pilot however commendable its ultimate objectives;

[27] A world benchmark has nothing to do with world totals. It is no more than what used to be done for the census of population — a call to arms for a purpose in a particular year. The notion is compatible with regional, subregional or interregional comparisons.

(c) Because there was no analysis of the differences between alternative aggregations, which confirmed the arbitrariness of the process.

118. This matter has received a good deal of discussion, both theoretical and applied. In fact, there is an implicit agreement embodied in the 1993 SNA about the way that figures should be compiled and disseminated:

> "The GK [Geary-Khamis] and the EKS [Èltetö-Köves-Szulc] methods have the same kinds of advantages and disadvantages as fixed price volume indices and chain volume indices in a time series context. The EKS index may provide the best possible transitive measure for a single aggregate between a pair of countries, in much the same way as a chain Fisher index may provide the best possible measure of the movement of a single aggregate over time ... The GK method is better suited to structural analyses of this kind [analyses that require information on the relative share of resources devoted to particular purposes in different countries or analyses that involve differences in relative prices] ... In general, the methods used to compile statistics must be influenced by the purposes for which they are to be used. As in the case of time series of national accounts, *it is therefore suggested that two sets of data should be compiled and published:*
>
> (a) EKS indices should be compiled for GDP and the main expenditure aggregates ... These would consist of both volume and PPP indices;
>
> (b) GK results should also be published in the form of values at the average prices of the block of countries expressed in some common currency, such as the US dollar."[28]

119. There is no aggregator function that combines the properties deemed desirable for all possible applications. Moreover, in the case of time-bound comparisons, time provides a natural order, and the comparisons of interest are usually restricted to consecutive periods and to comparison of all periods, with a base, arbitrarily selected but usually at the start of the run. In the case of space, there is no natural order and any comparison may be as legitimate and as interesting as any other. For this reason, lack of transitivity is much more of a handicap in the case of spatial than temporal comparisons. Be that as it may, the statement included in the 1993 SNA reflects a well-considered balance, and rather than revising it, it should be tried out systematically and effort be put into an analysis of the differences when these seem to be of material consequence.

120. The problem of the arbitrariness of ordering in space was discussed by Szulc.[29] The solution proposed is neatly stated in the following quotation from his paper: "... chain indices may be considered superior to their direct counterparts when they provide a smooth passage between the base and target time, rather than a detour". For time read space, and what remains is to formulate an acceptable criterion to choose the best passage possible. The conjecture advanced by Szulc is that this can be found by defining "distance" between countries, and when comparing Ucronia with Utopia choosing that chain which minimizes the distances between the two.[30]

---

[28] United Nations, *System of National Accounts, 1993* ..., paras. 16.102 and 16.103; italics added.

[29] See Szulc, op. cit.

[30] Szulc took as "distance" the sum of the absolute differences between two distributions rather than the Euclidean distance.

Bridge countries

121. It is well known that an attempt to link region A to region B by using country χ as the bridge produces results that are different from those obtained had country δ been used as a bridge. That too creates problems of credibility. There are two approaches — not mutually exclusive — that can be taken to surmount them. The first is to invest heavily in the parallel exercise that the bridge country's NSO must conduct. The second is to multiply the number of bridges with the double purpose of (a) examining how robust each single bridge is, and (b) if they are not robust, assembling sufficient data to use averages as a means of attenuating the idiosyncrasies of a single bridge country.

Bridge country

A bridge country, as its name suggests, spans two regions, that is, it can be used to compare PPPs estimated in region 1 with their counterparts estimated in region 2. In this respect, the bridge country is very much like a link year in ordinary time-bound index numbers.

122. The role of the bridge country is critical when we start to link regions with or without the intention of estimating a world GDP. For this reason, it is important to find a country that fits convincingly the role of bridge. There are several pairs of countries that appear to be very well suited for this purpose. For example, Mexico and Argentina compared with Spain are natural bridges for Latin American comparisons with Europe. The Szulc proposal would serve to answer the question on how to compare two poles, such as Bolivia with Iceland, and the constraint would be to make all paths go through say Mexico and Spain.

123. For this to work effectively, not only must some criterion, such as that of minimum distances, be upheld but the bridge countries must agree to the provision of two sets of data.

Proposal for a faster way of producing results

124. With current resources, know-how and capability of mobilizing all countries outside the OECD area, there is no question of attempting to produce a benchmark every year. But it is not acceptable to live with a situation in which official results appear less often than twice every decade, and then only with a lag of three or more years. It is also not satisfactory to do away with the notion of benchmark in much the same way as we do not cast out the notion of a census of population (or its administrative equivalent), to be completely replaced by a stream of small-scale sample surveys.

125. The *Handbook* rather optimistically mentions that "... benchmark estimates are not available until at least two or three years after the benchmark year", after noting that "... Typically, benchmark estimates are obtained every five years."[31] Neither of these statements is true any longer, but even assuming that in the best of cases this were the actual state of affairs, the results would still be of dubious relevance to any but the most obscure policy application. Worse still, if there were any important policy or operational application,

[31] See United Nations, *Handbook of the International Comparison Programme* ..., p. 62.

those responsible would have to invent an extrapolation method so as to respond to current concerns.

126. The *Handbook* goes on to note that "... The European Community has gone further in this direction, moving towards annual benchmarks. For EC, this partly reflects the fact that operational uses of real output numbers often require very current estimates". Apart from being a statement of the obvious, the *Handbook* describes a simple method of extrapolation, and rather unhelpfully concludes by noting that "... At present, there is not a recommended practice, and the method used is likely to depend on the specific purpose for which the extrapolation is carried out".[32]

127. There are different ways of running benchmarks. One is to conduct a five-year benchmark for all regions simultaneously. In order to keep this discussion simple, it is best to set OECD beyond the scope of the proposal, on the grounds that in any case the periodicity of its activities will be different. Moreover, once the PPP is stabilized there is a good chance that non-EU member countries will oblige OECD with an annual update of the benchmark estimate.

128. The alternative to a simultaneous benchmark is a rolling one, in which say one or two regions are handled thoroughly every year. Either way, information for a year will be provided by extrapolation, and extrapolated results will be amended once the benchmark results are compiled. The key point in this proposal is to accept the principle that information will be published on a preliminary basis and will be subsequently revised.

129. As is usually the case, there are two principal ways of handling national extrapolations. One uses the growth rates in the GDP components expressed at constant prices. The other consists in extrapolating PPPs by implicit price deflators. The second is probably the best forecast of the results expected from the new benchmark, although it does not preserve the national growth rates in GDP.[33] In a research mode, it would be very useful to try a variety of extrapolation techniques but it would not be helpful to publish a range of alternative preliminary figures.

130. If one wants to give ICP a higher profile its estimates must be used for a handful of key operational and policy applications, and that requires that they be produced quickly. The only way of coming up with credible results is to use all the information available from exchange rates, CPIs, implicit price deflators for each of the expenditure components, and if at all possible, mini-price surveys to supplement national information. The benchmark estimates should be accorded their proper role, which is to provide expanded detail and correct the extrapolations where they are found to be wrong. And in addition, there should be a programme designed to **learn from past mistakes**, meaning to improve the quality of the forecasts by the use of econometric techniques, plus any external information that may have a bearing on coming up with quick PPP estimates. This is an additional task for researchers and one that if successful may yield a very high rate of return.

A question of geography

131. The "geography" used in ICP is administrative. To the extent that the United Nations set up its economic commissions using economic and geographic criteria for so doing, those

[32] Ibid., p. 63.

[33] In handling this matter one had to be guided by the context in which the information will be used; for example, if the likely question will be "why are there two growth rates in GDP?", it is best to stick to the former method.

criteria are embodied in the regional aggregates estimated by the Programme. But to the extent that different rates of economic development have made the regions spanned by the economic commissions heterogeneous, the criteria have become obsolete. Some of the criticisms addressed to the Programme have to do with apparent difficulties in comparing goods and services provided by countries with very different institutional backgrounds and stages of economic development. Certainly, the ESCAP region is an example of acute heterogeneity. But so are the other regions.

132. There are two distinct problems in the aggregation of countries to a regional total. The first has to do with the dispersion of their GDP per capita, which in turn is closely related to the similarity of their expenditure patterns. The second has to do with the similarity of their tastes and institutions, which in turn affects the intercountry comparability of the goods and services they produce. The criticisms are about a perceived failure to deal with these two obstacles to international comparisons.

133. Table 1 provides an example of the relationship between GDP (or household consumption) per capita and the similarity in expenditure patterns. The example is taken from phase two of ICP,[34] and consists in comparing PPP-adjusted GDP per capita for six European countries with the corresponding figures for six developing countries in Asia and Africa,[35] and also in comparing the average index of similarity[36] for quantities consumed within the group of developed countries with the index of similarity between the group of developed and developing countries.

Table 1
Index of similarity within and between two groups of selected countries, and how it compares with their GDP per capita

(Phase two of ICP, data for 1973)

	Average index of similarity			
Countries	*Developed*	*Developing*	*Average GDP*	*Corresponding values of GDP per capita*
Developed	942	710	100	12
Developing	710	850	21	43

134. One way of deflecting the criticism and at the same time increasing the credibility of the Programme would be to calculate subregional aggregates and chain them explicitly in order to make comparisons involving remoter "poles".[37] In choosing subregional aggregates, special attention would be given to institutional, climatic and income per capita similarities. But there is no intimation that these criteria ought to be quantified and used

[34] See Kravis et al, op. cit., chap. 6.

[35] Belgium, France, Germany, Italy, the Netherlands and United Kingdom of Great Britain and Northern Ireland; and India, the Islamic Republic of Iran, Kenya, Malaysia, the Republic of Korea and the Philippines.

[36] The index of similarity is defined in Kravis et al, op. cit., chap. 6, and is no more than the correlation coefficient of the expenditure structures for any pair of countries. In this particular calculation, 34 expenditure groupings were considered. If the full 150 basic categories had been taken into account, the average indexes would have fallen very substantially, but this would not change the substance of the present argument.

[37] See Szulc, op. cit.

as weights. Naturally, this course of action would bring with it a set of new problems. Not least among them is the problem of a sparser set of data for any one aggregation.

135. A suggestion is set out below for what a classification of countries or areas by "distance" might look like. The proposal only includes countries or areas in Asia and in Africa, and makes no claim to completeness or to limit itself exclusively to those countries that have taken or are likely to take part in forthcoming phases of ICP. The criteria are obviously mixed. There is a geographic criterion, although some mega-countries span more than one of the subregions. There is an economic criterion. It is used in bringing together such countries as Malaysia and Indonesia. There are linguistic and cultural criteria — used for separating the West African countries into anglophone and francophone countries. No criterion is used exclusively, and there is no necessary consistency. China and South Africa stand on their own because of their uniqueness — relative size and relative stage of development, respectively. Japan, the Republic of Korea and the entrepôts have a difficult time being classified, and could easily be rolled into a single category distinguishable because of its GDP per capita.

Proposed classification of subregional groupings of countries or areas

Asia

(1) Countries bordering the Persian Gulf

(2) Indian subcontinent and adjacent countries (Afghanistan, Bangladesh, Bhutan, India, Myanmar, Nepal, Pakistan, Sri Lanka)

(3) Countries on the Indo-Chinese peninsula, excluding Malaysia, Thailand and Singapore

(4) Indonesia, Malaysia, Thailand, Philippines

(5) China

(6) Japan, Republic of Korea

(7) Hong Kong, China; Macao; Singapore

Africa

(1) Countries North of the Sahara (Algeria, Egypt, Libya, Morocco, Tunisia)

(2) East African countries (Burundi, Djibouti, Eritrea, Ethiopia, Kenya, Rwanda, Somalia, Uganda, United Republic of Tanzania)

(3) Southern African countries (Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, Swaziland, Zambia, Zimbabwe)

(4) South Africa

(5) "Desert" countries (Chad, Mali, Mauritania, Niger, Sudan)

(6) West African countries (Benin, Burkina Faso, Cameroon, Côte d'Ivoire, Gabon, Guinea, Guinea-Bissau, Senegal, Togo (francophone); Gambia, Ghana, Liberia, Nigeria (anglophone))

(7) Eastern island countries or areas (Comoros, Madagascar, Mauritius, Réunion, Seychelles)

(8) Central African countries (Central African Republic, Congo, Democratic Republic of the Congo)

136. In the case of the Economic Commission for Latin America and the Caribbean (ECLAC) region, after separation of the Caribbean from the mainland and singling out Central from South America (adding Cuba and the Dominican Republic to the mainland but leaving Haiti with the other islands), additional criteria might be economic. For example, the member countries of the Southern Core Common Market would form one bloc and the signatories of the Pacto de Cartagena another.

137. There are two objectives that this proposal seeks to attain. One is to maximize the number of instances in which a product drawn from a list is recognizable to the members of a subregion, either as a characteristic or as a near-characteristic product. The other is to meet such criticisms as those in the paper of the Indian Ministry of Industry.[38]

X. Steps to be taken in the short term

Three necessary features

138. There are at least three requirements that should be put in place in the reasonably near future in order to raise the Programme's credibility. In no particular order, they are a help desk; a set of explicit editing guidelines; and a small but vigorous and competent analytical (as opposed to research) capability. These three do not exist at present. The purpose of creating them is to improve the Programme's transparency (help desk and editing guidelines); its reach and impact (analytical capability); and its general credibility with users and producers alike (all three measures).

A help desk

139. The help desk is no more than an electronic network that allows someone at the entry point to put the questioner in touch with the best available expertise on whatever caused the question. For example, if the expert in Utopia is concerned that prices for commodity a may be out of line with those reported by his counterparts in country Ucronia, there is a telephone number or an e-mail address (preferably the latter) that will allow this kind of concern to be routed to the right place. The reason why these inquiries should go through a clearing house rather than being dealt with bilaterally is so that there is a trace of all concerns manifested during the collection and editing stages of a phase. Right now there is none. But without documentation there is no systematic feedback for subsequent phases.

140. The help desk must provide for a number of concerns. Three are readily distinguishable.

(a) A price compiler may wish to get help with an editing problem. One of the elements missing is to find out how average prices for a particular commodity in his country compare with the corresponding prices in neighbouring States;

(b) A price compiler may wish to get a ruling on whether a given deviation from the agreed list of goods and services, or the taking of a price in an unusual circumstance or the reduction in the agreed number of price quotes will be accepted. The integrity of the Programme demands that the ruling be documented;

[38] See Ghosh and Das, op. cit.

(c) A price compiler finds that a specific result involving his country does not appear reasonable but seeks guidance to find out if given the incidence of the problem on others supplementary efforts to solve the problem are required.

141. The staff manning the desk is therefore charged with three responsibilities:

(a) Routing the questioner to the place of best expertise, which implies keeping an up-to-date list of experts classified by competence so that prompt referral can take place;

(b) Recording the decision so that there is a documented trail, and keeping track of the questions asked in a form such that a subsequent audit is not impeded by difficulties of access or faulty memory;

(c) Checking whether the particular situation has already been the subject of a ruling, and if so whether the precedent can be used.

142. A summary of the questions generated by each phase should provide essential analytical and operational feedback to coordinators of subsequent phases.

Editing guidelines

143. The ICP *Handbook* is not useful on this matter beyond noting that:

> "... One principle of item selection that has been generally accepted, if not precisely defined, is that specifications priced by a given country should be sufficiently typical (characteristic) for the country. Pricing uncharacteristic items (i.e., goods or services which, though they exist in a given country, are not important in expenditure budgets and/or are not available in outlets for such items), is to be avoided. Items not commonly consumed may have very high prices etc." and that:
>
> "... The principle of choosing important products often comes into conflict with the second principle of choosing identical products. This is perhaps the most important issue in price selection."[39]

Unfortunately, the *Handbook* has very little else to offer on this "most important" issue, but a good deal else is required, particularly for NSOs without much experience on how to carry out the practical tasks of ICP.

144. And yet it is at the editing stage that international organizations are going to interact more intimately with NSOs. Leave aside those edits that are the result of improper data entry or the consequence of mistakes incurred by the price takers. For those errors, it is necessary that NSOs be at the same time committed and equipped with suitable computer-assisted techniques. Rather, this section addresses those cases where it is not possible for the national compiler to act without knowing what other countries in his region have reported. In fact, the interesting edits arise in situations where there was either improper specification for the particular good or service being priced, or alternatively there was an unworkable compromise struck between comparability and characteristicity. Necessarily, this is the kind of edit that can only result from data confrontation before the data are aggregated.

145. The regional coordinators should play two roles in this regard. First, they must display these cases in a clear and convincing fashion. For this, they need the means in terms of computer software and a way of getting in touch with their counterparts at NSOs. The editing

tables drawn up in appendix III* are one form of display. No doubt they can be improved upon once this role is accepted. Second, the coordinators may have to broker a change leading to a respecification of the good or service, or else to a new compromise between characteristicity and comparability.

146. In order to act credibly in any one of those two roles, explicit editing rules must be drawn up and agreed to by those responsible in NSOs. For example, one could say that all list items for which the among country dispersion of prices is of at least two coefficients of variation on either side of the mean should be questioned. The understanding is that a question might lead to a revision to the agreed specification or else a revision to detailed country submissions. As far as the regional coordinators are concerned, they would have to learn how to coordinate decisions among themselves, which is possible if the role of the United Nations Statistics Division in this matter is reaffirmed and embodied in the person of a world coordinator.

147. Implicit in these remarks is the need to post agreements on editing and treatment to be given to the list of goods and services as a phase is launched. These agreements can be refined in the course of the data-collection stages and indeed one of the roles of the help desk should be to assist in refining them. Thought should also be given to the possibility of making available current news on rejected outliers in the form of a notice on an electronic bulletin board.

An analytical capability

148. An ICP phase provides innumerable opportunities for research work in an academic setting. There is no doubt about that. The question is whether a similar number of opportunities is provided to policy analysts. Clearly, until such time as the results are available considerably faster, the question should hardly be posed. So for the purposes of this discussion we must assume that timeliness is a solved problem, and the question is how to present in an analytical framework the results of a new phase.

149. The results of an ICP phase include information on G-K and E-K-S aggregations, on volume and price indices, on per capita GDP and its expenditure components, on expenditure shares in regional totals, on extreme cases that may distort the overall picture and so on. In addition, at the end of the sixth phase, in spite of all the incomparabilities owing to different numbers of participants — the consequences of variable political geography — there is a unique stock of data capital that has been accumulated and can help users to understand the sense of current numbers.

150. There are only two sets of circumstances that justify not including with the publication of a large body of new data, a reasoned attempt to answer at least two questions:

(a) What did the results of this survey or of this statistical construct show that we did not know from earlier episodes or from related data?

(b) Why is the knowledge conveyed not trivial? *or* Why is it important that I be aware of the new information?

---

* The appendices to the present report are available for consultation in its electronic version, at: www.un.org/Depts/unsd.

[39] See United Nations, *Handbook of the International Comparison Programme* ..., p. 30, paras. 115 and 116.

Either the answer to those questions can be found elsewhere or else speed in conveying the information is of the essence, and in any case it is information that fits naturally into a well established framework.[40] Examples of what is meant by analytical accompaniment can be found in the United States Bureau of Economic Analysis issues of the *Survey of Current Business*, which include the publication of the latest quarterly GDP numbers.

151. Where neither of the two sets of circumstances mentioned above prevails, the release of the data is the first and perhaps the only occasion to establish their importance with potential users. But that cannot be done by lengthy explanations of the methods by which the data were derived, the classifications used, the standards that guided data collection in the more complex cases and so on. While these are necessary adjuncts to any professional publication of the numbers, it is to the questions listed above that answers must be provided. The need to provide such answers is all the greater when the lag between the period of reference and the date of publication is significant.

152. But in spite of the need and the abundance of data that have to be analysed, the analytical content of the texts accompanying releases is meagre. For example, two recent publications on purchasing power parities in the Economic and Social Commission for Western Asia (ESCWA) and African regions[41] devote most of their textual material to a description of methods and standards. One of them adds a few pages describing in words the results of a handful of scatter diagrams. In neither case is there an attempt to show by example how the data can be used to inform important decisions and how its alternatives could give rise to less fortunate choices. Since in the better of those two cases three years elapsed before the data were published, a rush to publication is hardly a convincing explanation for the lack of analytical content.

153. The argument that G-K and E-K-S aggregations provide very different results and there is no established way of analysing the differences is not very convincing either. Assuming that the convention of publishing both aggregations continues as suggested in the 1993 SNA, it is difficult to imagine policy applications for which the difference at the level of broad aggregates would affect the nature of the advice.[42]

154. The type of analysis that should accompany a release ought to include the following elements:

(a) How does the ranking of country within the region differ from the ranking if market exchange rates have been used, and what are the principal factors that explain the differences?

(b) How does the current situation differ from the situation described in the previous phase, both in terms of the evolution of PPP-adjusted aggregates and in terms of the difference between those aggregates and the ones estimated using prevailing market exchange rates?

---

[40] One could argue that the publication of the monthly CPI must be as prompt as possible and most analysts can draw appropriate conclusions given the supporting data for the overall index.

[41] United Nations, *Purchasing Power Parities: Volume and Price Level Comparisons for the Middle East, 1993* (United Nations publication, Sales No. 97.II.L.7); and *Comparison of Price Levels and Economic Aggregates 1993: The Results of 22 African Countries* (Luxembourg, 1996).

[42] An experiment using per capita GDP and gross fixed capital formation numbers corrected for PPP for both the ESCWA region and the 22 African countries that took part in phase six shows correlations of the order of .98 (in the case of ESCWA, excluding Palestine). So long as the differences in scale are adjusted, the comparisons do not involve more than the countries within one single region, and do not go below the very broad aggregates, the differences in aggregator function do not matter much.

(c) Having taken into account possible biases and errors in reporting, together with the degree to which data have had to be imputed, how far off from the "truth" are the estimated figures likely to be?

(d) How has the structure of expenditures in the region changed since the last phase, and what redistributions have taken place within the region?

The final analytical chapter would be produced once estimates for all regions are in and it can be ascertained how the region has fared in the world's pecking order since the last phase.

EuroCost's publication on the African countries

The purpose of singling out this publication for critical comment is not because it stands out as especially worthy of criticism but because it is as good an illustration as any of the relative neglect with which the stage of dissemination of ICP results has been treated. The fact that publications have been of indifferent quality has not contributed to increase the credibility of the Programme in the eyes of its critics.

The publication has 32 pages of introductory text, followed by 22 pages of tables (English version). The first 15 pages are factual and definitional, and include four pages devoted to alternative aggregation methods — not enough pages for a thorough discussion and for too many for a cursory reference. The next 17 pages are so-called analytical but fail to raise issues of interest to either international or national agencies, particularly as they appear three years after the year of reference.[a] The triviality of some of the findings borders on the comic. For example, accompanying chart 3 on page 37, there is a statement that Egypt's volume index of production of alcoholic beverages is comparatively low. True but hardly worth mentioning as the finding of a world programme designed to estimate purchasing power parities.

On the other hand, there are no references to difficulties that countries may have experienced, instances of non-compliance with the list of specifications and references to extreme cases that may or may not have been dropped from the calculation of binary indices. There is no reference to how many imputations were required and how confident EuroCost is of their quality and so on.

---

[a] The matter of analytical text is quite fundamental in restoring credibility to the exercise. Having descriptive text that virtually reproduces in words what a few scattered tables and graphs convey is of little consequence. Until the analytical text can show a finding of which users should take notice, there is no virtue in delaying publications for the sake of including it.

XI. Cost, recommendations and conclusions

How much will it cost?

155. There is little sense in adopting a bureaucratic posture by adding up the costs and submitting to the treasuries of the various interested organizations a request for the corresponding finance. The Programme has not built up a stock of credibility sufficient to justify such an approach. On the other hand, the current state of financial affairs will not allow urgent measures to be taken, in particular to prepare for a new phase with new objectives and in many ways a broader scope. The advice is to seek inspiration in Simon Goldberg's[43] approach to the financing formula for the National Household Survey Capability Programme. There is a difference though between the two programmes. NHSCP was essentially a framework. The objectives were to be fixed by NSOs. In the case of ICP, the output is known and well defined. The financing formula aims at systematizing the Programme, making the process describable and explicit, and restoring to it the credibility that is required if the Programme's outputs are to be regarded as useful.

156. The financing proposal is based on the assumption that the programme will be structured in the way described in table 2.

Table 2
Structure of proposed programme and purpose of expenditures, by financing entity

Benchmarks	*International organizations*	*NSOs*
	Seminars; travel; world coordinator; regional coordinators; seed money for selected NSOs; consultants; dissemination	Technical staff directly provided to the coordinators of the Programme; supplements to CPI and other price programmes
Current updating	Coordinators; data collection; travel and communications	Data collection and provision; communications

The benchmark exercise can be conducted in one of two ways: (a) the classical way, in which there is a single reference year and the aim is to calculate an adjusted world GDP for the year in question; (b) the alternative way, which is to conduct rolling benchmarks in which the reference year varies from one region to the next over say a five-year cycle. Table 3 provides an example of what the alternative scheme could look like:

[43] For those whose contact with international statistical matters is recent, Simon Goldberg retired from his position as Deputy Chief Statistician of Canada to become the fourth Director of the then United Nations Statistics Division. And after retiring from that position, he devoted his last few active years to the creation, institution and resourcing of the Survey Capability Programme by building a consortium of interested partners who took over the financing of the project.

Table 3
Hypothetical benchmark cycle

Region	*Organization*	*Year of 5-year cycle*
Asia and non-OECD Oceania	ESCAP	One
West Asia	ESCWA	Two
Africa[a]	ECA	Three
Caribbean	ECLAC	Four
Central and South America (mainland)	ECLAC	Five

[a] The assumption is that EuroCost is not involved in future phases; obviously, this is a worst case assumption.

157. The following are the assumptions supporting the cost estimates: the OECD region looks after itself and most of it is on an annual schedule anyway; the salaries of the technical staff are paid for by sponsoring NSOs; every year there will be three seminars, two for the region in scope and one in preparation for the next region; some of the analysis will be contracted out to expert consultants. One further assumption is that for the first cycle, the only prices collected are for household expenditures.

158. With these assumptions and using the costs incurred by the Latin American project as a basis, annual direct costs would be of the order of:

Table 4
Annual direct costs for a five-year benchmark cycle
Thousands of United States dollars

Purpose of expenditure	*Amount*
Seminars	150
Travel and communications	150
Seed money[a]	150
Coordinator salaries[b]	350
Consultancy contracts[c]	50
Total	**850**

[a] Money required to help selected NSOs to conduct supplementary price surveys etc.
[b] Including the salary of the world coordinator and supplements to the costs of part-time regional coordinators.
[c] Mostly to help analyse new data and to update benchmark information.

Table 4 excludes expenses by NSOs, which imply the use of their staff on price collection and editing; sponsorships by NSOs; and salaries already paid to part-time coordinators. It also excludes the expense of the World Bank's research agenda.

159. The point about these very rough indications is to suggest that worldwide the effort can be conducted at a cost of less than one million dollars annually, which should not be beyond the reach of a consortium of interested organizations. In any case, it would be the first responsibility of the world coordinator to test the waters and determine what are the prospects for financing on this scale.

Recommendations

160. The following are the recommendations of this report:

1. ICP should not be ended nor should it be allowed to languish. Indeed, if no adequate financial support can be found for it, languishing would be its worst fate.

2. Securing financing on a broader scale implies making a commitment to producing reliable and timely data, with well documented methods and sound analytical commentary.

3. While the long-term objective remains that of estimating all the components of final demand, in a first stage the compilers should have the modesty of making do with price estimates for household expenditures.

4. The savings accruing from a more restricted price collection should be ploughed right back into the Programme.

5. The Programme must have a global or world coordinator.

6. The coordinator must be known, respected and with demonstrated administrative and professional abilities (the expression "professional" involves a grasp of the complex of national accounting, economic applications and basic statistics) to coordinate a project of this size and complexity.

7. A new phase of the project must start with a resolution endorsed by the Statistical Commission. That resolution should follow the submission of a document that sets out unambiguously what is expected, why it is being done, what means will be used, what are the responsibilities and accountabilities of the participants, and what are the standards of quality that are aimed at in connection with ICP.

8. A Commission resolution should be the result and not the cause of mobilization of resources. The latter should be preparatory to final approval.

9. The next phase should be designed in such a way as to produce a continuous information based either on a benchmark study or else updated through the help of consumer price indexes and exchange rates.

10. Additional resources to the project should be obtained by creating training positions attached to the coordinator and financed by sponsoring statistical offices.[44]

11. The dissemination activity must be shared between NSOs and the statistical arms of international agencies. Country statistical offices must be enlisted in order to give the project greater visibility and a stronger sense of relevance.

12. Interested parties (the United Nations, IMF, the World Bank, the Asian and Inter-American Development Banks, selected NSOs) should mobilize the required resources under the guidance of the world coordinator to:

(a) Prepare an ICP prospectus as a means for discussion with prospective contributors;

[44] The modalities of this proposal would have to be worked on and be subject to administrative and financial constraints imposed on the one hand by the United Nations and on the other by the sponsoring countries. But permanent stationing in New York, Washington, D.C., Paris or Luxembourg is no longer necessary so long as there is a good communications network that will allow for video conferencing and intensive exchange of views via e-mail, phone and fax.

(b) Recruit the first set of short-term assistants after defining the modalities of their contract;

(c) Embark on the drafting of a revised and expanded handbook (probably manual is a better term) offering guidance to NSOs engaged in ICP.

Conclusions

161. No statistical programme with an international dimension needs central coordination and an effective relationship with NSOs more than ICP. The soundness of the Programme requires that both national and international offices play their role effectively. As a result it is more vulnerable than average to personality conflicts, small changes in budget, apparent lack of direction and so on. On the other hand, a strong hand, a feeling of commitment and purpose and the rallying of support on the part of NSOs can turn the situation around quickly. A programme evaluation is an opportunity to ask fundamental questions about the wisdom of keeping a programme alive. If the answer is "yes" and the cost can be made affordable, the other decisions to ensure survival should not be too difficult to take.